

A PRIORI KNOWLEDGE

by

Markos Valaris

B. Sc. National Technical University of Athens, 2001

Submitted to the Graduate Faculty of
Philosophy in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

University of Pittsburgh

2008

UNIVERSITY OF PITTSBURGH

SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Markos Valaris

It was defended on

April 30th, 2008

and approved by

Cian Dorr, Associate Professor of Philosophy, University of Pittsburgh

Michael Thompson, Professor of Philosophy, University of Pittsburgh

Sebastian Rödl, Professor of Philosophy, University of Basel (Switzerland)

Dissertation Advisor: John McDowell, University Professor of Philosophy, University of Pittsburgh

Copyright © by Markos Valaris

2008

A PRIORI KNOWLEDGE

Markos Valaris, PhD

University of Pittsburgh, 2008

The goal of my dissertation is to give an account of our capacity for a priori knowledge in terms of fundamental features of conceptual thought. In a nutshell, the claim is that a priori knowledge is possible because thinkers are able to recognize the norms that their thinking is subject to. But, as I argue, the *contents* of our propositional attitudes are themselves individuated in terms of the norms that fix their place in our thinking. Since contents — Fregean thoughts — are what may be the case, it follows that knowledge of the norms of thinking can be substantive knowledge of the possible shapes reality can take.

TABLE OF CONTENTS

PREFACE	VII
INTRODUCTION.....	1
1 THE AIM OF THE PROJECT.....	1
2 OVERVIEW OF THE COMING CHAPTERS	2
CHAPTER 1: THINKING AND THOUGHTS.....	9
1.1 INTRODUCTION	9
1.2 RUSSELL AND LOGICAL ATOMISM.....	13
1.3 FREGE.....	20
CHAPTER 2: CONCEPT-POSSESSION AND A PRIORI KNOWLEDGE	35
2.1 INTRODUCTION	35
2.2 THE ANALYTICITY STRATEGY	36
2.3 A HOLISTIC CONSTRAINT ON CONCEPT-POSSESSION	52
CHAPTER 3: TWO-DIMENSIONALISM AND THE EPISTEMOLOGY OF RECOGNITION.....	64
3.1 INTRODUCTION	64
3.2 THE TWO-DIMENSIONALIST PICTURE OF CONCEPTUAL THOUGHT	65
3.3 RECOGNITION AND INFERENCE	70

CHAPTER 4: BASIC A PRIORI KNOWLEDGE AND KNOWLEDGE OF ONE'S OWN MIND.....	89
4.1 INTRODUCTION	89
4.2 THE QUESTION OF JUSTIFICATION.....	91
2.3 OVERVIEW	97
2.3 BASIC A PRIORI KNOWLEDGE AND SELF-AWARENESS.....	105
APPENDIX A: ON THE INDIVIDUATION OF FREGEAN THOUGHTS	118
1 INTRODUCTION	118
2 A LINGUISTIC APPROACH.....	118
3 AN ALTERNATIVE APPROACH	124
4 UNDERSTANDING AND PROPOSITIONAL ATTITUDES	131
BIBLIOGRAPHY	143

PREFACE

Many people gave me help and support essential to my being able to write this dissertation, in many different ways and in many different capacities. The list that follows will have to be incomplete, both in that it cannot mention them all, but also in that it can hardly begin to express my gratitude towards those that it does mention.

First of all, I am grateful to John McDowell, my dissertation advisor, whose intellectual guidance is reflected in every single page of this work. He took the time to carefully read and comment on every bit of this dissertation, many times over. Just as important, however, was his personal support during the early stages of my project — when it seemed to me simply unfeasible. Without it there is no doubt that this dissertation would not have been written at all.

But every single member of my committee contributed in invaluable ways to my project, and I am grateful to them all. Cian Dorr and Sebastian Rödl both read several drafts of vital parts of this dissertation, and helped me with insightful and challenging comments. Their influence is obvious in many places, as is that of Michael Thompson's ideas.

My way of thinking about philosophy in general, and about my project in particular, owes a lot to conversations I have had over the years with many people around the department. Most obviously, the members of my committee have been crucial in this regard. But there are others too who are owed special thanks here: conversations with Tom Ricketts, Mark Wilson, Lissa Merritt, Sasha Newton and Greg Gates have been exceptionally important.

My passage through graduate school would have been much harder, or even impossible, had I not had the enormous good fortune to enjoy the friendship and kindness of many people, some of whom I have already listed. First among them is Lissa Merritt — my wife. Her patience and support — even when not living together — not only saw me through many rough periods, but also taught me to be more patient and supportive of myself. But I have throughout also relied on the support of people far away: my family and friends back in Athens, as well as in Paris, Hong Kong, Berlin, London and elsewhere. The little time I get to spend with them each year means much more to me than I can possibly express here.

INTRODUCTION

1. THE AIM OF THE PROJECT

We have knowledge of the world from many different sources and of many different kinds. Most obviously, a lot of our knowledge about the world can be traced back to some particular causal interaction between us and things in our environment: we know things through perceiving them, for example, or through remembering having perceived them, or through other people's telling us about them. Regardless of the details of the epistemology of perception, memory, or testimony, the fact that we can acquire knowledge in these ways is naturally taken to point to a deep fact about us: whatever else might be true of the human mind, it is the sort of thing that is open to the causal influence of the world in a special way — radically unlike, say, the way in which a stone or even a plant is open to the causal influence of the world.

It is not, however, evident that all of our knowledge can be traced back to such causal interactions — not, at least, in the relevant sense. Even though for each one of us there was a particular process through which we learned enough basic arithmetic to know that $5+7=12$, for instance, it would appear that our *grounds* for that knowledge do not consist in the authority of the person or the book responsible for this learning (even if, plausibly, the learning was necessary for the real grounds of that knowledge to become available to us).

Moreover, it is also arguable that the first kind of knowledge — the knowledge that we have through experience, or our causal openness to the world — would be either impoverished beyond

recognition or even altogether impossible, if we did not also possess abilities of a sort explicable only by the attribution of knowledge which (on pain of circularity) would have to be grounded otherwise. Even to pick out an object in one's visual field as "that thing over there", for instance, one arguably needs to have some knowledge of the identity conditions of spatially extended objects, and so on. This is not to deny that experience may, again, play an enabling role in having this kind of knowledge. But if there is indeed some knowledge that has to be in place in order for one to possess the most basic ability for cognitively significant experience, then it is hard to see how such knowledge could itself be *grounded* in experience.¹

Any purported example of the relevant kind of non-empirical knowledge is bound to be controversial. It is not the aim of this dissertation — much less this introduction — to sort out particular such controversies. I am assuming that, despite the controversies, there is an intuitive case to be made for some attributions of such knowledge — to give it a name, *a priori* knowledge.² The aim of this dissertation is to address the following question: what facts about us would the possibility of *a priori* knowledge point to?

2. OVERVIEW OF THE COMING CHAPTERS

1. If *a priori* knowledge is not grounded in interactions with the world, then how is it grounded?

One traditional answer — the one I will defend in this dissertation — is that it is grounded in the

¹ This kind of argument has its origin in Kant, and it is characteristic of the Kantian tradition (Kant (1998), Strawson (1999; 2002), Evans (1982; 2002a)).

² I cannot here demarcate the kind of knowledge that is the topic of this dissertation otherwise than just by alluding to purported examples and to the negative characterization "non-empirical". However, I do not think that this negative characterization suffices to pick out a principled epistemological category, and the examples — as I already admitted — may be contested. This seems inevitable at the outset of a project such as this. The positive account I arrive at by the end should help demarcate a genuine kind of knowledge.

nature of thinking itself. This is the view Kant develops at great length, and also the view Frege expresses in the following famous pronouncement:

What if beings were even found whose laws of thought directly contradicted our own, and therefore frequently led to contrary results even in practice? I would say: here we have a hitherto unknown kind of madness.³

Inability to recognize the laws of thought — that is, for Frege, at least the laws of logic and (therefore) arithmetic — must be a sign of “madness”, of defective mentality. Contraposing, it is a requirement of *non*-defective mentality that one be able to recognize the laws of thought.

Now, it seems clear that, if we wish to maintain that there is a connection between a priori knowledge — and knowledge of logic in particular — and non-defective mentality, we will have to qualify Frege’s pronouncement in some way: even assuming classical logic to be correct, for example, it is not at all obvious that non-classical logicians must either suffer from some form of madness, or else fail to reject the classical laws of logic they take themselves to be rejecting. I discuss such points in Chapter 2. But, before getting to this, there seems to be something more immediately surprising about the very idea that a priori knowledge is grounded in the nature of thought — for why should it be that the nature of thought is a good guide to the way the world is? Why should even *non*-defective mentality entail any access to facts about the world, other than the access to the world that consists in the mind’s openness to causal influence?

As I argue in Chapter 1, Russell for one found this idea unintelligible, except perhaps within the context of an objectionably idealistic framework. Russell’s commitment to rejecting internal relations in general, and — more specifically — internal relations between thought and its objects, led him to reject the Kantian and Fregean approach to the epistemology of the a priori as well.⁴ For, Russell seems to have thought, how could the nature of thought provide access to facts about the

³ Frege (1997: 203). The passage makes more sense if we imagine scare-quotes around “laws of thought” in the first sentence.

⁴ Russell (1997: 86-89).

world, unless the world were somehow bound up with the nature of thought? For this reason, Russell prefers to postulate a separate faculty of a priori intuition, over and above our capacity for thought itself.⁵ On Russell's view, we might say, inability to recognize the truths of logic would be a kind of *blindness*, rather than a kind of *madness* — it would be, that is, a cognitive defect compatible with a perfectly developed capacity for thought.⁶

Russell's positive account of a priori knowledge is not satisfying, not least because he does not have anything to say about that supposed faculty of a priori intuition. But the question he raises is, I think, a legitimate and important one: how are we to understand the relation between thinking and its objects, if a priori knowledge is to be grounded in the nature of thought?

Naturally, the question is pressing only on the assumption that there really is such a thing as a priori knowledge of facts about the world. There are various views, ranging from old-fashioned conventionalism to contemporary non-factualism, which may be seen as attempting to neutralize Russell's question by denying precisely this. According to conventionalism, the subject-matter of a priori knowledge is not the world, but, rather, the way *we* represent the world in thought or language. If we can make sense of this suggestion, then Russell's question loses its bite — since there would be nothing surprising in taking the way we represent the world not to be independent of the nature of our minds. According to non-factualism, a priori disciplines — such as logic and mathematics — have no subject matter at all: they merely express some of the dispositions that collectively make up our cognitive or linguistic faculty. Again, if we can make sense of this suggestion, the effect would be to neutralize Russell's question — since to say that logic and mathematics have their source in the nature of our minds in this sense does not require that our minds be related to any objects at all.⁷

⁵ A priori intuition, in Russell's sense, is of course not related to Kantian a priori intuition.

⁶ The analogy with blindness was suggested to me by Cian Dorr.

⁷ For a historical discussion of conventionalism, see Coffa (1993). For a non-factualist view, see (for example) Field (2001). There is also, of course, the radical empiricism of Quine and others, which rejects the idea of a priori knowledge altogether. I briefly address some aspects of this position in the Appendix, but

I do not directly discuss conventionalist or non-factualist views in this dissertation. On the contrary, in Chapter 1 — and in the Appendix that supplements it — I offer a straight answer to Russell's question. In particular, I argue for the Fregean view that thoughts — that is, the objects of propositional attitudes — are individuated in terms of the norms of thinking that govern individual subjects (or even entire practices, such as — to use Frege's own example — arithmetic). But thoughts are what may be the case; thus, knowledge of the norms that govern our thinking is, at the same time, knowledge of the possible shapes reality can take. Insofar as we can explain our capacity to know the norms that govern our thinking, therefore, there is (in principle) no difficulty in taking this to be a capacity for knowledge about the world.

2. The idea that a priori knowledge is to be explained in terms of the conditions of individuation of thoughts is currently a popular one. But it is an idea that needs to be handled with care. In Chapters 2, 3 and 4 I criticize recent accounts of a priori knowledge which — in one way or another — start out from this idea, and I propose an alternative.

More specifically, Chapter 2 criticizes an assumption about concept-possession that is the starting-point of many contemporary authors. This is the assumption that possessing a concept must be a matter of possessing a fixed set of dispositions to judge or infer in accordance with certain rules, which are constitutive of the concept in question. A subject's capacity for a priori knowledge, in turn, is to be explained in terms of such dispositions.

In fact, it often seems to be taken for granted that a Fregean approach to a priori knowledge — that is, an approach that ties the individuation of contents to the norms of thinking, and a priori knowledge to knowledge of the norms of thinking — requires this assumption (Timothy Williamson, for example, in attacking this assumption takes himself to be attacking the Fregean

otherwise it does not figure in this dissertation. This is not simply a matter of economy of resources: the Quinean position is motivated by a conception of the questions of epistemology and philosophy of mind that is quite alien to the one adopted here, so that a common ground for a fair discussion would be hard to find.

approach in general).⁸ But, as I argue in that chapter, this is not correct. The Fregean strategy can — and *should* — be pursued without this assumption.

This is not to say that possessing a concept makes no difference to the thinking of the subject who possesses the concept, and neither is it to say that the difference it makes is not crucial in understanding our capacity for a priori knowledge. But, as I argue in Chapter 2, we should not think of concept-possession atomistically. As I suggest, to possess a concept is a matter of participating in a conceptual practice. Membership in such a practice, in turn, does not require that one conform perfectly to all of the norms constitutive of the practice, but only that one be reliable in conforming to *enough* of the norms of the practice. Even deeply entrenched and widespread a priori mistakes are possible on this view, as long as we can make sense of the subject as a member of the relevant conceptual practice.

Chapter 3 expands the criticism of existing views of conceptual thought along the same lines, but it is to some extent parenthetical. More specifically, in this chapter I criticize another recent attempt to carry out a version of the Fregean strategy, namely, the two-dimensionalist approach of Frank Jackson and especially David Chalmers. Although the two-dimensionalist view also makes the assumption I criticize in Chapter 2, and so would be subject to the objections discussed there anyway, it also seems to be crucially mistaken about a further aspect of conceptual thought — in particular, it seems to require a very implausible view of the recognitional capacities we bring to bear in applying concepts in experience. But the point of the chapter is not just negative: although the nature of the recognitional capacities involved in experience is not a central topic in this dissertation, the discussion helps round out the picture of conceptual thought that underlies my account of a priori knowledge.

⁸ Williamson (2006).

In the fourth and final chapter of this dissertation I complete my account of the nature of our capacity for a priori knowledge. The holistic constraint on concept-possession proposed in Chapter 2 helps explain why we are, in general, reliable in conforming to enough of the norms governing our conceptual practices; but it does not, so far, explain why that conformity counts as a capacity for *knowledge*. For all that has been argued so far, a subject might possess the relevant reliable dispositions and yet still be in the dark about why she is disposed to conform to the patterns she actually conforms to, or even about what patterns she conforms to. Such “blind” dispositions would not, I take it, count as expressions of a capacity to know. As I argue, however, they are not even recognizable as the kind of disposition that constitutes our capacity for conceptual thought at all; thus they do not pose a threat to the idea that a priori knowledge is grounded in the nature of conceptual thought.

The problem here, I believe, is that we have so far failed to take into account the self-awareness characteristic of conceptual thought. This failure is another important shortcoming of recent accounts of a priori knowledge. As I explain in Chapter 4, judgment and belief are essentially “reflexive” — that is, they essentially involve a kind of self-awareness. In this respect, our cognitive capacities are radically unlike the blind dispositions described above. Combining this point with the holistic constraint on concept-possession argued for in Chapter 2, I argue that our most basic capacity for a priori knowledge consists in our reflexive ability to reliably conform our thinking to enough of the norms constitutive of our conceptual practice.

This finally completes my account of how a priori knowledge is possible. But to what extent is the end product an epistemology of the a priori?

I certainly make no attempt to formulate a set of necessary and sufficient conditions for a state of a subject to count as a state of a priori knowledge. This is not what this dissertation

purports to offer. However, I think that this is not a reasonable request to make: as has forcefully been pointed out, the prospects of giving a set of necessary and sufficient conditions for knowledge — hence, *a fortiori*, for a priori knowledge — are extremely dim.⁹ Attributions of mental states in general, and of states of knowledge in particular, are made in the context of attempting to make sense of human subjects and their interactions with their environment. We should not expect to be able to capture the conditions for such attributions in a set of informative necessary and sufficient conditions.

Giving a set of necessary and sufficient conditions for a priori knowledge, however, is not the only intelligible goal of an epistemology of the a priori. I am assuming here that, even lacking an analysis of the concept of knowledge, we still have enough of a grip on it to be capable of consistent and confident enough attributions of knowledge, and even of making distinctions among different kinds and sources of knowledge. As I pointed out at the beginning of this introduction, we can take these intuitive judgments and distinctions as starting points of an investigation into the nature of the mind. What I try to offer in this dissertation is an account of aspects of our capacity for thought that make some of our intuitive attributions of knowledge intelligible.

⁹ For example, McDowell (1998a), Williamson (2000).

Chapter 1
THINKING AND THOUGHTS

1. INTRODUCTION

1. The overarching goal of this dissertation is to develop an account of our capacity for a priori knowledge. It is natural to think that, if such an account is possible, then it must be intimately related to an account of thinking. If that were not so — if our capacity for a priori knowledge, that is, were simply independent of our faculty of thought or understanding — then it would appear to be a pretty mysterious capacity indeed. Perhaps that appearance of mystery is no more than *just* an appearance — I have no explicit argument to rule out the possibility that an account of a priori knowledge in terms of an independent faculty could be given. Nevertheless, this dissertation will only focus on the prospects of grounding an epistemology of the a priori on an account of thinking.

This chapter will focus on sketching an account of thought with a view to preparing the ground for the more explicitly epistemological tasks undertaken in the rest of the dissertation. Broadly speaking, there are two main questions this dissertation aims to address. On the one hand, if there is any plausibility to the idea that a priori knowledge is grounded in the nature of thought, then that must be because there is some intuitive plausibility to the idea that being a thinker gives one some sort of epistemic access to facts about the structure of thought. I think this is true; but we need to get clear on which facts these are, and why they should be accessible to thinking subjects just in virtue of their capacity to think. This task is undertaken in Chapters 2 to 4. The present chapter addresses a more fundamental question about the nature of thought.

Even granted that some account can be given of why being a thinker should entail some sort of access to facts about the structure of thought, why should we take such knowledge to be anything other than just knowledge of facts about *us*, and the way *we* happen to represent the world? Why should we, in other words, take it to be substantive knowledge of mind-independent reality? My fundamental goal in this chapter is to suggest a view of thinking and its objects that addresses this question (the suggestion is refined and supplemented in the Appendix to this dissertation).

For reasons that will become clearer as the argument proceeds, I approach the issue through a discussion of two different conceptions of thinking and its objects which, historically, have shaped the debate in the analytic tradition. These are the views of the logical atomists and especially Russell, on the one hand, and of Frege on the other.

2. What is the relationship between the nature of the objects of thought (and the semantic contents of sayings) and the nature of the activity of thinking?¹ According to both Frege and the logical atomists, whatever that relation turns out to be, it is crucial to avoid a reduction of the objects of thought to subjective items, items that properly belong to the subject-matter of empirical psychology. According to both, the main point is to “always separate sharply the psychological from the logical”, to use Frege’s own phrase.² One might find in the Kantian tradition still prevalent in Frege’s time a tendency to assume that an account of the nature of the *acts* of thinking is at the same time an account of *thoughts*. From the perspective of Frege, however, the assumption that the two accounts can be simply collapsed rests on some sort of act/object ambiguity: the *objects* of thinking — “thoughts”, in one sense of the term, or *Gedanken* in Frege’s German terminology — are not

¹ I will throughout simply assume that the contents of sayings and the objects of intentional states or acts are things of the same sort. I will use the verb ‘to think’ as the most generic term for a mental relation between a subject and an object of the appropriate sort: thus to entertain, to judge, to question, to hope are all in the terminology here adopted kinds of thinking.

² Frege (1980: x). This is the first of the three maxims Frege announces in the introduction to the book will be guiding his work. For my conception of the form in which the Kantian tradition was handed down to Moore and Russell I rely on Hylton (2002).

themselves mental entities, although our *acts* of thinking them are dateable events taking place in each person's individual mental life.

The anti-psychologistic stance adopted by Frege is quite intuitive. That more than one person can have the same thought at the same or at different times seems to be a truism; but, far from being a truism, it would actually be trivially false if it entailed a multiplicity of occurrences of the same dateable and localizable mental event. It is therefore reasonable to explicitly draw a distinction where the tradition Frege was responding to did not (or, at least, did not seem to Frege to do so). The obvious way to draw such a distinction is to introduce terminology to talk about entities distinct from the dateable events that are our acts of thinking, and which serve as their objects. Thus Moore and Russell talk about propositions, and Frege talks about thoughts.³

Now this distinction, even at this completely abstract level, appears to raise two questions. The first question concerns the nature of the relation between these new entities and our minds. To say that these entities are the objects of our acts of thinking is, at best, to give a suggestive metaphor, and substantive work may be required in order to get beyond that. The second question concerns the relation between thought and the world. The proposed entities are what we think when we think and what we express in language when we speak. But one of the essential characteristics of thought and speech is that they are often intended to be responsible to the way things are. How are we to account for the responsibility of thought and speech to the world — that is, how are we to account for the possibility of truth and falsehood — on this picture?

³ See, for example, Russell (1996) and Frege (1997b; c). The motivation I mention in the text is clearly visible in Frege (1997c). Frege (1997b) is perhaps more famous for putting thoughts to use in Frege's account of indirect contexts. Dealing with such contexts, however, means confronting the problem posed by statements such as 'I believe the same thing as Jones', or 'I always believe what Jones says'. Thus indirect contexts provide just another way in which the non-psychological nature of the objects of thinking and the contents of sayings emerges.

I will ignore throughout worries about the ontological status of abstract entities in general.

Frege and the logical atomists diverge in their respective responses to these questions in quite crucial ways. In what follows I want to trace some of these divergences.

In §2 I discuss the distinctive approach to these questions developed by the logical atomists, and especially Russell. Russell famously dismisses the idea that a priori knowledge can be grounded in an account of our capacity for thought. This is how he puts it, while discussing the apriority of the law of contradiction:

The law of contradiction is about things, and not merely about thoughts [...]. If this, which we believe when we believe the law of contradiction were not true of the things in the world, the fact that we were compelled to *think* it true would not save the law of contradiction from being false; and this shows that the law is not a law of *thought*.⁴

As I will suggest, Russell's dismissal of the possibility of grounding an epistemology of the a priori in an account of the "laws of thought" is very closely connected to certain central features of the logical atomist framework (features that survive as late as *The Problems of Philosophy*, where the quote comes from).

In §3 I will sketch the alternative, Fregean view. The Fregean view, in contrast to the Russellian one, makes it intelligible that a capacity to know facts about the structure of our own capacity for thought — the "laws of thought" — could in principle also be a capacity to know facts about the world. Frege himself exploited this possibility in his own work on logic and the foundations of mathematics — work that is in large measure epistemological, and not only metaphysical.

⁴ Russell (1997: 89).

2. RUSSELL AND LOGICAL ATOMISM

1. According to the well-known view of the logical atomists, a proposition is a special kind of complex of simple entities.⁵ These simple entities are conceived of atomistically, that is, as completely independent of each other. The constituents of propositions include, in Russell's terminology, "things" and "concepts". These two categories are distinguished by the role items belonging to each of them can play in propositional complexes: things can only occur in a proposition non-predicatively, while concepts can occur either predicatively or non-predicatively. The structure of propositions is analogous to the grammatical structure of sentences. Thus, an atomic proposition will consist of one or more items — either things or concepts — in non-predicative occurrence joined by a concept functioning predicatively, to yield a whole capable of truth or falsehood. For instance, to the sentence "Desdemona loves Othello" corresponds a complex consisting of Desdemona and Othello, joined by the relation of love.⁶ It is natural to speak of an object that figures only non-predicatively in a proposition as what the proposition is *about*; and this is an expression I will sometimes use in what follows.

According to logical atomism, therefore, the fundamental constituents of propositions are at the same time the fundamental constituents of reality: particular things, properties, and relations. This opens up the prospect of dealing straightforwardly with the second of the two problems mentioned near the end of the previous section: there is no ontological gulf between the things that we think or say and what is the case in the world. Propositions, if true, just are what the case is. Regardless of the details of a coherent account of truth consistent with logical atomism, the present

⁵ The account of the view I am presenting here derives mostly from Russell (1996), as amended by the elimination of non-propositional complexity by means of the theory of descriptions presented in Russell (1905) (this is what allows me to straightforwardly say that propositions are composed of simples). I will not consider the changes in Russell's views that culminate in the multiple-relation theory of judgment, and the concomitant excision of propositions from his ontology. The early Russell's conception of proposition has, in any case, survived its author's change of mind and plays an important role in contemporary debate.

⁶ Russell (1996: 45-48).

account of the nature of propositions makes it intelligible that propositions are the sort of thing that gets a grip on reality.⁷ This might make logical atomism seem like an attractive view about the objects of thinking. There are, however, problems with it that I believe show that it is deeply flawed.

Logical atomism gives a clear theoretical priority to the mind-independent constituents of reality. In accounting for the nature of propositions Russell relies on a prior, and perhaps intuitive, understanding of the basic ontology of the world. Propositions are then understood as a special kind of composite object, built up in a special way from particular objects, properties and relations. This is Russell's way of ensuring a consistent anti-psychologism: there is nothing due to the mind in the proposition as Russell conceives of it.⁸ But this characterization of the nature of propositions obviously makes pressing the first of the two problems I raised at the end of the last section: how are minds related to things of that sort?

In a sense, if it has no satisfactory answer to this question logical atomism only *appears* to avoid the traditional difficulties concerning the possibility of thought's being related to reality. By construing the objects of thought as built up from the constituents of reality, logical atomism only succeeds in pushing the problem back to the question of what it is for a proposition to be the object of an act of thinking. As I will argue, no satisfactory solution is available within the framework of

⁷ My formulation in the text has to be cagey, because Russell himself did not think it an easy matter to give a precise and technically adequate account of truth within logical atomism (Russell (1996: 48-49)). It is a bit hard to see exactly what Russell's problem is, but it clearly arises from the combination of the view that facts are true propositions with the view that truth is a property of propositions. The combination of these two views certainly has results that Russell may have found difficult to swallow. From this combination it follows that the circumstance of *p*'s having the property of being true consists in the truth of the proposition that *p* is true, while the truth of that proposition consists in the truth of the proposition that it is true that *p* is true — and so on. The problem here is not really the infinity of the ensuing regress, since that is arguably not vicious; it is, rather, that this account does not make explicit the *nature* of truth (see Hylton (2002: 178-179); Ricketts (2001: 105)). If this is the problem Russell has in mind, then there are obvious ways around it — one could, in fact, simply drop the demand for further explanation at this point. As a matter of historical fact, Ricketts argues that this problem played a major role in Russell's own later abandonment of the ontology of propositions — but I am not here going to consider this change in Russell's views.

⁸ See, for example, Russell (1906).

logical atomism. In particular, it is *atomism* that makes the problem so acute: if one assumes, as Russell does, that simple things cannot stand in internal relations to each other, and complex things can stand in such relations only insofar as they share some of their constituents, it is very difficult to make sense of the idea that propositions are the objects of acts of thinking.⁹

2. To think a proposition is to stand in an epistemic relation to it. The one constraint Russell throughout his career imposes on this relation is encapsulated in his famous principle of acquaintance: in order for a proposition to be available to a subject to think, the subject must be acquainted with all of the proposition's constituents.¹⁰ The principle of acquaintance itself states only a necessary condition for thinkability; but it is quite clear that Russell would be willing to regard acquaintance with all of its constituents as a sufficient condition of being able to think a proposition as well.¹¹

Thus Russell's notion of acquaintance, a notion notoriously hard to precisely pin down, lies at the center of his account of thinking. Russell never presents an analysis of the relation of acquaintance: it is, according to him, a primitive relation, which admits of no further analysis. But he does provide a theoretical constraint on it, as well as a number of examples he considers self-explanatory. The constraint is that acquaintance is a form of knowledge of *objects* (both particular things and concepts), independent of knowledge of *propositions*. Acquaintance is a form of immediate presence to the mind of an object:

⁹ It is often thought that logical atomism prohibits *any* kind of internal relation. Even though Moore and Russell may have been prone to use this slogan themselves, it seems clear that their real view must be closer to the one I present in the text. For example, a proposition essentially has the constituents that it has: if a proposition *p* is about an object *a*, then it is essentially about that object. By the same token, if another proposition *q* is also about *a*, then it too will be essentially about *a*. It seems to follow from these two premises that *p* and *q* essentially stand in the relation of being about the same thing.

¹⁰ Russell (1997: 58; 1905: 456-457).

¹¹ For an argument that Russell is committed to this see Levine (1998: 417-419).

We shall say that we have *acquaintance* with anything of which we are directly aware, without the intermediary of any process of inference or any knowledge of truths.¹²

It is not clear from this formulation how this independence from inference and from propositional knowledge should be understood. However, the background of Russell's thought shows that it has to be understood in very strong terms. In fact, as I will suggest, the constraint is so strong that it is quite implausible that there is any relation answering to Russell's conception of acquaintance — at least not one that could plausibly serve as a foundation of our capacity to think most of the thoughts that we take ourselves to think.

First of all, it is important to keep in mind that acquaintance is a form of *knowledge*. Thus Russell does not offer a causal theory of reference: even though causal relations might be necessary in order for a subject to be acquainted with a perceptible thing, they are surely not sufficient. It is not always the case that a state or act caused by a perceptible object can have as its content a proposition with that object as a constituent: in order for that to be possible, the subject must moreover *know* the object in an appropriate way. The question is how to spell out what the required sort of knowledge is.

The crucial point is that the notion of immediacy or directness that Russell employs in discussing this sort of knowledge has to be understood in especially strong terms. The context from which I extracted the quotation above may suggest that the contrast he has foremost in mind is that between a kind of non-inferential knowledge of an object and the kind of knowledge of an object one has in virtue of possessing a description uniquely satisfied by it. This contrast is certainly there, for the latter kind of knowledge evidently consists in knowledge of truths about the object, and may moreover require a process of inference to figure out which object satisfies the description. But in fact I believe Russell intends to state a thesis significantly stronger than that. I believe Russell is

¹² Russell (1997: 47-48).

committed to denying not only that acquaintance is equivalent to knowledge of an individuating description, but also that it relies *in any way* on propositional knowledge.¹³

The reason why Russell is committed to this view has to do with the way in which he understands the relations between the objects of thought, on the one hand, and minds on the other. Propositions and their constituents are, as we have seen, totally mind-independent items. Nevertheless, they are supposed to be the objects of mental acts. On a certain understanding of mind-independence, these two theses entail that, whatever it takes for an entity to be the object of a mental act, it must not entail any non-trivial condition on what that entity is. This view is explicitly stated by Moore:

[Propositions and their constituents] are possible objects of thought; but that is no definition of them. It merely states that they may come into relation with a thinker; and in order that they *may* do anything, they must already *be* something.¹⁴

The context makes it clear that by the term ‘definition’ Moore here means something like ‘real definition’, or an account of what something is; and his claim is that what propositions and their constituents are is independent of the fact that they are possible objects of thinking.

I suggest that it is this feature of logical atomism that leads Russell to maintain that acquaintance cannot have any propositional knowledge as a necessary condition. For, suppose that acquaintance with a certain object has propositional knowledge about that object as a necessary

¹³ One might be led to accept this stronger thesis by means of an inference from the principle of acquaintance. The principle of acquaintance requires that in order to understand a proposition *p* one must be acquainted with *p*’s constituents. If this is so, then it might appear problematic if it turned out that in order to be acquainted with *p*’s constituents one had to know *further* propositions about these objects — for in order to know these further propositions, one would once again have to be acquainted with the very same objects; but it was precisely acquaintance with these objects that we supposed required knowledge of these propositions in the first place. Thus, one might think, we are led into what looks like a potentially vicious circle. But in fact it is not at all clear that there is a vicious circularity here. If one already believed that thinking a proposition must, at the bottom level, rest on a non-propositional epistemic relation with its constituents, then it evidently follows that the circularity ought to be ruled out. But this is precisely the point under question here: why should we agree that thinking a proposition requires a prior, non-propositional epistemic relation with its constituents? If one rejects that assumption, then taking our fundamental epistemic relation to objects to consist in knowing some propositions about them does not seem problematic at all.

¹⁴ Moore (1899: 179).

condition. Then it would seem that one could know some facts about an object, not by collecting evidence about how things happen to stand with that object (after all, in order to do that, one would already need to be acquainted with the object), but simply on the basis of one's ability to think about it. But then, it seems that the mere fact that the object is a possible object of thought has non-trivial consequences about its nature — for otherwise how could we come to know things about it simply by reflecting on our own ability to think? The possibility of such knowledge, therefore, seems to violate the atomist thesis articulated in the passage from Moore above.

This is not to say that logical atomism has to deny that, as a matter of contingent fact, our minds are such that we have to think in accordance with certain patterns. What is not compatible with logical atomism is only that there are patterns that could constitute a source of propositional knowledge about the objects of thought — that there are, as Russell puts it, “laws of thought”.¹⁵ To admit the possibility of such knowledge would appear to undermine the Russellian conception of the mind-independence of propositions and their constituents, for such knowledge, instead of tracking the way things are, seems to anticipate it. Acquaintance must not presuppose any propositional knowledge: it must be, in this sense, presuppositionless.¹⁶

3. The problem, of course, is that such a presuppositionless way of relating to the constituents of the propositions one thinks about is hard to make sense of. Merely causal contact would be presuppositionless in the relevant sense; but, as we have seen, Russell's acquaintance is more than just causal contact — it is a form of knowledge. At the very least, it would seem that being acquainted with an object in Russell's sense requires being able to single it out. But it seems clear

¹⁵ Russell (1997: 86-89).

¹⁶ It should be noted here that Russell's descriptivism does make room for a kind of knowledge that anticipates, rather than tracks, certain facts about objects. But, of course, on the present reading of Russell this is one way of explaining why knowledge by description cannot be the fundamental way we relate epistemically to objects of thought.

that — at least in the most salient, intuitive cases — this is not a non-propositional, presuppositionless ability.

Consider, for example, the case of perceptual reference. It is natural to think that one can think of an object in virtue of focusing one's visual attention on it — picking it out, at the very least, as *that thing over there*. But it is also natural to think that the ability to pick out things by means of perceptual demonstratives is not the sort of presuppositionless ability it would have to be to satisfy Russell's conception of acquaintance: in picking out something by means of a perceptual demonstrative, one plausibly needs to bring to bear a prior conception of spatially extended things that persist through time.¹⁷

Thus, from this point of view, it is not surprising that when Russell gives examples of the things we are acquainted with, he does not talk of such things: the things he includes within the domain of acquaintance are sense-data and other introspectible items like memories, and universals he supposes we can abstract from those.¹⁸ If there were *anything* we can know in the required presuppositionless way, then we should expect to know *these* things in this way.

Of course, it is in fact quite doubtful whether even introspectible items of this sort can be known in the way Russell supposes: it would appear that even in order to pick out a sensation that one has, for example, one needs to bring to bear a prior conception of a subjective history constituted by a sequence of distinct events. But even if we do not press this objection for the moment, it is clear that Russell's conception of acquaintance leads to very implausible consequences about our account of thinking. What I have tried to show in the last few sections is that Russell's conception of acquaintance is itself closely related to fundamental features of the logical atomist framework. As I will suggest in what follows, an alternative account of thinking, which is not

¹⁷ See Evans (1982: ch. 5).

¹⁸ Russell (1997: 47-52). The abstractionist story about universals is given in pages 101-105.

committed to that framework, is both more plausible and more congenial to the project of giving an account of a priori knowledge.

3. FREGE

1. Now, there may be ways to fix the logical atomist picture in order to avoid the problems sketched above. The more recent idea of direct reference, for instance, may perhaps provide a way to cut through them. Theories of direct reference weaken the characteristic Russellian principle that our attitudes towards propositions have a foundation in an epistemic relation between us and the constituents of these propositions, while retaining Russell's conception of propositions as constructed out of ontological materials independent of the nature of thinking.¹⁹ The form of contact with its constituents required for thinking a proposition is, on these views, not subject to the exceedingly strong constraints that Russell's notion of acquaintance was.

What I am interested in here, however, is not so much a *fix* of the Russellian picture, but rather the significantly different view Frege develops on the same issues. According to Frege, the objects of thought are independent of all individual minds. Nevertheless, this does not entail that being an object of thought is intelligible independently of the principles of rationality that determine what it is to be a thinker: the nature of the objects of thought and the nature of the act of thinking are not independent of each other.

Frege's account of thoughts — the objects of acts of thinking — as the senses of sentences is meant to capture precisely this idea. In the first instance, the question that puzzled Frege had to

¹⁹ See, for instance Kaplan (1989). Such views also draw on some of the themes in Kripke (1999). In light of their abandonment of Russell's strong notion of acquaintance, the relation between such views and Russell's is not as close as it might be assumed. Such views are typically developed with *linguistic* reference foremost in mind, generally addressing questions about thought less directly than Russell does. It is, therefore, not always clear how they would respond to Russell's motivation for his notion of acquaintance.

do with identity sentences: how can a subject, without any failure of understanding or rationality, assent to “ $a=a$ ” while refusing to assent to “ $a=b$ ”, even though a and b are one and the same object? To the extent that ‘ a ’ and ‘ b ’ are genuinely referring terms, and not definite descriptions, it seems clear that Russell would have difficulties accommodating this case. But it seems evident that cases like this must be possible, for sentences of the form “ $a=b$ ” often express significant empirical discoveries.²⁰ According to Frege, this phenomenon is possible just in case “ $a=a$ ” and “ $a=b$ ” can be used to express different thoughts — or, equivalently, have different *senses*. Generalizing the point, we can say that for Frege a fundamental principle is the following (in a formulation adapted from Evans):²¹

The Intuitive Criterion of Difference: A thought T is different from the thought T' if it is possible for someone to understand both at a given time while coherently taking mutually exclusive attitudes towards them, i.e. accepting one while rejecting, or being agnostic about, the other.

Some qualifications are required. For one thing, the criterion as stated does not address the possibility of thoughts that cannot be thought by one subject at the same time, but not because of constraints of rationality. First-personal thoughts provide one such example for Frege. I will not address such cases in this dissertation.

Moreover, it is not in general to be expected that in each case the identity of the thought a subject is thinking will be transparent to her: a subject may not be in a position to recognize a thought as the same again on different occasions of its being occurrently present in her consciousness, even assuming no intervening lapses of memory, lack of attention, or other failures of this sort on her part. We are often able to make much better overall sense of a subject’s total

²⁰ Frege (1997b: 151). To be precise, the question Frege asks in this essay is this: how can learning the truth of an identity sentence one already understands be a cognitive achievement over and above the derivation of a trivial consequence of the law of identity?

²¹ Evans (1982: 21). This principle is explicitly appealed to by Frege in a number of places. See, for example, Frege (1997b: 156; 1997c: 332-333; d: 128).

psychological state by admitting the opacity to her of the objects of some of her attitudes.²² But if one is in such a situation, then one might rationally hold mutually exclusive attitudes towards the same thought, simply because one does not know it is the same thought. Thus ordinary understanding cannot in general be sufficient for our criterion, for such understanding is intuitively compatible with failure to re-identify a thought.

Further, the Intuitive Criterion as here stated only offers a sufficient condition of difference, or, equivalently, a necessary condition of sameness. The general shape of a necessary condition of difference is, I think, quite clear: one has to look into the constituent structure of thoughts, and require that thoughts that are distinct must differ somewhere in that structure.

I offer some more detailed suggestions concerning these two issues in the Appendix to this dissertation. What is important, for now, is that this approach to the individuation of the objects of thought marks a crucial difference between Frege and Russell. For one thing, Fregean thoughts evidently are more finely individuated than Russellian propositions. But this is just a symptom of an even deeper difference. For Russell, as we saw, the objectivity and the non-psychological nature of propositions were supposed to be secured by taking them to have in general no necessary relation to anything mental. For Frege this is evidently not the case: the principles of identity and distinctness of thoughts are intimately connected with the principles of rationality we use (together with pragmatic principles governing communication) in order to make sense of each other. For him the objectivity and the non-psychological nature of thoughts are, rather, secured by the fact that the principles of rationality are not empirical generalizations, but principles concerning what it is to have a mind at all. This is how he puts it, in a striking formulation:

²² Cf. Evans (1982: 21). The possibility of such cases (and Frege's apparent willingness to embrace them: see Frege (1979)) contradicts a thesis traditionally attributed to Frege, according to which sense is transparent, in being necessarily re-identifiable in just the way denied in the text. See, for example, Dummett (1975: 131); Levine (1998: 433). A notable exception is Burge (1990). I discuss this point further in Appendix A.

Logic [does not have] the task of investigating minds and the contents of consciousness owned by individual men. [Its] task could perhaps be represented rather as the investigation of *the* mind; of *the* mind, not of minds.²³

It is not of course to be expected that the principles of individuation of thoughts, which are at the same time the principles governing “*the* mind”, will be immediately available to any individual thinker. Perhaps some thinkers are not even capable of acquiring explicit knowledge of these principles. This should not, however, detract from the basic point: the nature of mind and the nature of the contents of thinking are not independent of each other in the way in which they were for Russell.

This raises, now, the question of how contents, as conceived by Frege, are supposed to relate to the world — it raises, that is, the question of the possibility of truth and falsehood. I discuss this in the next section. In the sections that follow I discuss how the Fregean conception of content forms the basis for his approach to a priori knowledge.

2. Thus Fregean thoughts are individuated in a way that is sensitive to the considerations of rationality that guide our practice of psychological attributions. This by itself, as we saw, is enough to contradict Russell’s atomistic conception of the mind-independence of propositions. But the tight connection between the nature of Fregean thoughts and principles of rationality is not an unargued-for assumption: it is rather the product of an even more fundamental characterization of thoughts, as those items “for which the question of truth and falsity can arise at all”.²⁴ This point needs some spelling out, because it seems on the face of it that the question of truth and falsehood, and the extensional logic that Frege claims provides the “laws of truth”, do not require the fineness of grain that the principles of individuation for thoughts provide us with. In the context of Frege’s

²³ Frege (1997c: 342).

²⁴ Frege (1997c: 327-328). Frege says that he does not intend this as a *definition* of thought, that is, as a non-circular account of what thought is. This, as I will suggest later on in the text, is because Frege does not think that truth itself has a definition in this sense. We can perhaps give an explanation of what truth is, but that explanation will depend on our grasp of what a thought is.

logic all we are interested in is the relation of an object's falling under a concept; thus it can make no difference to that logic *how* one refers to that object or concept. Getting clear on what the connection is between Frege's conception of thoughts as essentially truth-evaluable and the fine-grained principle of individuation he adopts for them is quite crucial for my purposes in this chapter, because (as will become evident as the argument proceeds) it is here that the possibility of a priori knowledge begins to take shape.

What connection is there, then, between the question of truth and the fine-grained distinctions that questions of rationality bring in?

The point here is, in a nutshell, that for Frege the concept of truth belongs in a tight circle of concepts, including those of thought, judgment and assertion, whose relations to each other are in part constituted by the principles of rationality that give rise to the fineness of grain that questions of truth by themselves do not seem to require. These relations are encapsulated in Frege's characterization of thought as that concerning which the question of truth arises, of the act of judgment as the act of acknowledging a thought as true, of assertion as the manifestation of a judgment.²⁵ That the concepts of thought, judgment and assertion are connected through principles of rationality is I think clear from the discussion in the previous section. Nevertheless, one might think that truth is a different matter, since the concept of truth intuitively relates the intentional or the linguistic with the extra-intentional and the extra-linguistic. In fact, insofar as the concept of truth bears connections to the concepts of thought, judgment and assertion it might be hoped that an independent account of the concept of truth would allow us to crack open the circle these concepts form.

This, however, is not at all how Frege thinks of the matter: according to him, there is no account of the nature of truth to be given — for example, as a kind of correspondence with reality

²⁵ Cf. Ricketts (1986: 71). My discussion here is heavily indebted to this paper and to Ricketts (1996) — although I diverge from his interpretation in some important respects (as noted below).

— other than what we get by its interrelations with the other concepts in the circle. For Frege, any attempt at formulating non-circular necessary and sufficient conditions for truth — a *definition*, in Frege’s strict sense of the term — is bound to fail. From this he concludes that “the content of the word true is *sui generis* and indefinable”.²⁶ It is not to my purpose here to examine in detail Frege’s arguments for this view; I will, however, try to indicate what sorts of reasons one might have for holding it.

For Frege, as we saw, a thought is that for which the question of truth arises. To entertain a thought is to consider whether things are thus-and-so. To judge a thought is to take the *thought to be true*. But this only means that to judge a thought is to take *things to be* thus-and-so. Taking a thought to be true — that is, judging a thought — is not a further ‘taking’ with a thought as its subject matter, over and above ordinary takings whose subject-matter (for the most part) consists in the things around us. This seems to be Frege’s point in the following passage:

It is something worth thinking about that we cannot recognize a property of a thing without at the same time finding the thought *this thing has this property* to be true. [...] It is also worth noticing that the sentence “I smell the scent of violets” has just the

²⁶ Frege (1997c: 327). This is a place where I diverge from Ricketts (1986). According to Ricketts, the proper conclusion to draw from Frege’s argument (together with his conception of judgment) is that the expression ‘is true’ is not a predicate of thoughts at all, in the sense that it does not signify a Fregean concept. Ricketts focuses on Frege’s regress argument, according to which defining truth in terms of some other property or properties (say, as being *F*) would lead us to a vicious regress. The regress gets going, according to Frege, because since judgment is the recognition of the truth of a thought, in order to judge that *p* one would first have to determine whether the thought that *p* is *F*. But determining this will itself be a judgment — namely, the judgment that the thought that *p* is *F*. Thus we embark on a vicious regress.

Now, Ricketts claims that this argument does not require the detour through being *F* at all: the mere assumption that truth is a property of thoughts ensures that judging anything will involve judging that a thought has a certain property, and hence will get us going on the regress. I have no space here to explore this issue in the detail it deserves, but I will briefly indicate why I think Ricketts’s argument is not convincing. The reason is that Frege’s claim that judgment is the recognition of the truth of a thought cuts both ways. Thus, emphasizing one direction, it says that *recognizing the truth* of a thought is simply to *judge* that thought (Frege’s examples of redundant uses of ‘is true’ may be taken to illustrate precisely this point). If this is right, then the supposedly distinct cognitive act of recognizing a thought as true — the act that gets us going on the regress — will simply collapse into *judging* the thought. Thus the regress cannot get started. This point, I believe, effectively blocks any attempt to analytically define truth — for then we *would* have to determine separately whether a thought has the defining properties (see the main text for a more precise formulation of this point) — but it does not tell against the idea that truth is a *sui generis* property, which we can grasp only through our acquaintance with the concepts of judgment, assertion, and thought.

same content as the sentence “it is true that I smell the scent of violets”. [...] Nothing is added to the thought by my ascribing to it the property of truth.²⁷

Frege’s claim, now, is that this would not work unless the concept of truth is given only through its interrelations with the notions of judgment, thought, and so forth. If truth were to be identified with some independently specified property or set of properties, then it would be an open question whether taking a thought to be true — that is, to have this independently specified property or set of properties — could be just the same act as taking the things that are the subject-matter of the thought to be one way rather than another. For any independently specified property, it would be an open question whether taking a thought to have *it* would simply amount to judging the thought.²⁸

All this suggests that Frege has no need for coarse-grained entities such as Russellian propositions. One might perhaps think that entities of that sort, built up as they are supposed to be from building-blocks available independently of anything pertaining to thought, would be required in order to explain truth. As we saw, however, Frege does not think that truth admits of any such explanation. Of course to deny that truth admits of any such explanation is not to suggest that for Frege thought is in any way cut off from reality. Truth is to be understood together with concepts such as that of taking things to be thus-and-so, and specifically in terms of *successful* such takings. But takings are successful only if things really are as they are taken to be. It is, therefore, natural for Frege to simply *identify* true thoughts with facts, with what the case is.²⁹

²⁷ Frege (1997c: 328).

²⁸ Cf. Ricketts (1996: 77). One might legitimately question whether this “open question” style of argument really blocks all candidate reductions of truth, and in particular reductions that do not purport to give an analysis or definition of the concept. The question of such reductions, in the guise of a posteriori semantic naturalism, is very much a live issue. A proponent of such a project may agree that truth cannot be understood independently of the other concepts in the circle, but still hope that all of these concepts may turn out to refer to a system of physical properties (perhaps *via* a system of higher-level functional properties). I agree that the open question argument does not tell against this possibility. For Frege it was probably not in view at all. Thus everything I say here is strictly independent of the question of a posteriori reduction, insofar as such a reduction leaves in place the conceptual connections here outlined.

²⁹ Frege (1997c: 342). See also McDowell (2005: 56). It might seem that the fineness of grain this identification induces in the individuation of facts is counter-intuitive: how can the *fact* that Afla is at least 5000 meters high be distinct from the *fact* that Ateb is at least 5000 meters high? Yet it is clear that these

There is here a superficial similarity with Russell's view of propositions: Fregean thoughts, just like Russellian propositions, not only are truth-evaluable, but are — if true — facts, or what the case is. But the similarity is only superficial. For Russell this result is obtained because propositions are complexes whose components and unity have nothing to do with the mind. His claim that true propositions are facts, therefore, only serves to make problematic the accessibility of facts to mental acts such as judgment. For Frege, on the contrary, the identification of true thoughts with facts is the result of his interlocking understanding of the concepts of thought, judgment, truth and so forth. Facts, as true thoughts, are individuated in accordance with principles of rationality, with the principles that govern *the* mind — although, of course, they are independent of any individual mind, or even of there actually being any minds at all.

Frege's claim that thoughts belong in a "third realm", neither that of "things in the external world" nor that of "ideas",³⁰ is actually liable to obscure the real force of his own position here. Talk about a third realm distinct from that of ideas and that of things in the external world may suggest that the inhabitants of that realm are intermediaries of some sort, mediating our epistemic contact with the external world. But it is quite clear that Fregean thoughts are not in any sense intermediaries: they are the things we think, and also the things that, if true, are what the case is. They are of course not *material* things, because they are things only in a sense which includes abstract — *unwirklich* — entities as well as tables and stones; but the facts about concrete things in both the external world and the realm of ideas are true thoughts, and so inhabitants of the third realm.³¹

thoughts may be distinct by Frege's lights. Frege does not address this issue directly, but it should be remembered that, for many purposes, thoughts and their constituents matter only insofar as they determine *reference*. But since reference is coarsely individuated for Frege, it follows that for many purposes distinct but related thoughts — for example, thoughts related in the way that the thought that Afla is at least 5000 meters high is related to the thought that Ateb is at least 5000 meters high — can be lumped together. Such lumping together, however, does not bring us closer to anything more "real" than the thoughts we started out with.

³⁰ Frege (1997c: 337).

³¹ This is not to say that the "third realm" thoughts inhabit is to be identified with the realm of *abstracta*, such as redness or the number one, as suggested in Burge (1992: 300). Thoughts are significantly different from

Seeing the connection between rationality and truth in Frege's thinking suggests that requirements of reason can be a perfectly adequate source of knowledge about the world. From this perspective there are no grounds for denying the possibility of substantive a priori knowledge arising from the "laws of thought" — a view that for Russell was completely out of the question, as we saw above. Russell was thus forced to adopt a quasi-perceptual epistemology of a priori knowledge: all he could offer on that score was an insistence that, among the things we are acquainted with, there are relations between concepts, and that it is this acquaintance that enables us to know non-empirical propositions.³² On the Fregean view, on the contrary, there is in principle room for a quite different epistemology of the a priori: a priori knowledge can, on this view, be acquired through reflection on the requirements of thinking itself. In closing this chapter I want to give an initial sketch of what such an epistemology of a priori knowledge might look like.

3. The most straightforward way to approach Frege's epistemology of the a priori is through a brief consideration of his conception of logic. This, in turn, requires taking account of his views on the structure of thoughts and of the sentences that express them.

Fregean thoughts are individuated, one might say, from the top down, unlike Russellian propositions which are individuated in terms of their composition out of independently available

Fregean functions (including concepts) and abstract particulars such as the number one, precisely because thoughts are things that *may be the case*. Redness and the number one are among the things we think about; *what* we think when we think about them are thoughts, inhabitants of the third realm. In that respect functions and abstract particulars are closer to tables and stones than to Fregean thoughts, although — like thoughts — they lack spatiotemporal and causal properties. (Of course we can also think thoughts about thoughts; but that is not what is distinctive about them.)

Burge does not even consider the reading I am here proposing. He suggests a threefold classification of readings of Frege: one that construes Frege as an idealist who took facts to be the products of thinking, one that construes him as a deflationist who took metaphysical issues to be issues about language or thinking, and his own preferred Platonist reading (Burge (1992: 303-304)). My reading does not fit in this classification at all: for, unlike the Platonist, I *do* claim that there is a constitutive link between thinking and fact; but, unlike either the idealist or Burge's deflationist, I deny that this gives any priority, either metaphysical or analytic, to thinking. In fact I believe that my reading is better suited to the rationalist epistemology Burge finds in Frege than his own Platonist one — although of course I cannot argue for this here.

³² Russell (1997: 86-9). See also the discussion of "intuitive knowledge" in chapter XI.

building-blocks. This is not to say that Fregean thoughts are not essentially complex — as I already suggested (and as is argued more explicitly in Appendix A), we cannot give a full account of the individuation of thoughts without considering their internal structure. The crucial difference between Frege’s account of the composition of thoughts and Russell’s account of propositions is that, for Frege, the constituents of thoughts are not independent atoms, but essentially such as to combine in determinate ways. In *Foundations of Arithmetic* — that is, before explicitly developing the apparatus of sense and meaning, and hence before introducing his technical notion of a thought — Frege announces his famous context principle: “only in a [sentence] do words really have a meaning”.³³ It is natural to think that, once he developed his characteristic view of thoughts as the senses of sentences, he would have held the analogous thesis concerning thoughts and their constituents.

The context principle in the first instance only seems to concern sub-sentential expressions and their senses. But it should be noted that the principle should extend as well to compound sentences and thoughts, with other sentences or thoughts as constituents.³⁴ This extension could not of course be correctly expressed by saying that sentences express thoughts *only* in larger sentences. The point is rather that the possibilities of combination sentences or thoughts have are

³³ Frege (1980: §60). I have altered the translation by replacing the word ‘proposition’ by the word ‘sentence’, in order to avoid ambiguity. Approaching Fregean thoughts and sub-sentential senses from the top-down allows us to take a charitable approach to Frege’s assimilation of sentences to complex proper names. Whatever the formal similarities that motivated this view, it is widely acknowledged that from a philosophical perspective it is a bad idea, since it obscures the deep difference between the sense of a sentence (a thought, capable of truth and falsity) and that of an ordinary name. (See, for example, Dummett (1981: 196)). According to Ricketts (2002), this assimilation is the main confusion Wittgenstein accuses Frege of in the *Tractatus* and in the writings that precede it. Taking the top-down approach, however, allows one to recognize that sentential senses — thoughts — are fundamental, since these are the points where the connections with both rationality and reality (*via* truth) are made. The sense of sub-sentential expressions is to be understood as the contribution they make to the sense of the sentence. From this perspective, therefore, we can restrict Frege’s assimilation of sentences to names to something like an awkwardness in his formalism, which does an injustice to his own thinking.

³⁴ In fact something stronger is true on Frege’s official view. Construing sentences as names of truth values, he also construes sentential connectives as first level concept-expressions. Thus there is in Frege’s symbolism no distinction corresponding to the intuitive one between atomic and compound sentences. This is an awkward result of Frege’s formalism, and one that has naturally drawn fire from his readers.

essential to them: it is, for example, essential to any sentence p that it can combine with any other sentence q in a way such that they compose a new sentence, expressing a new thought, which is true if and only if the thought expressed by q is true or the thought expressed by p is false.³⁵ In the same way, Prior's pseudo-connective 'tonk' does not represent a way thoughts can combine with each other.³⁶ The laws of these combinations — the propositional part of logic — are then partly constitutive of Fregean thoughts.

Similarly, the quantificational part of logic spells out some of the internal structural possibilities of sentences and thoughts. The logical relations between a generalization and its instances, as well as those between sentences or thoughts that share sub-sentential components, correspond to the object/concept distinction Frege takes as fundamental in his analysis: what it is for a linguistic expression to stand for an object, or for a thought-component to be a mode of presentation of an object, is in part spelled out in the patterns of valid (truth-preserving) quantificational inference. The same is true, *mutatis mutandis*, for Fregean concepts and predicates, and the thought-components corresponding to them.³⁷

From all this, together with the thesis that facts are simply true thoughts, it follows that if one has a priori knowledge of logic one thereby has a priori knowledge concerning the nature of things. In that sense, then, logical knowledge is substantive knowledge. That Frege thought so, of course, is evident from his writings on logicism, and especially from the way in which he presents his

³⁵ Frege (1967: 13-4).

³⁶ 'Tonk' is defined in terms of its introduction and elimination rules, as follows: from P infer ' P tonk Q ', and from ' P tonk Q ' infer Q . It can easily be seen that there is no truth-function that can satisfy these rules.

³⁷ Notice that neither the claims of this paragraph nor those of the previous one provide a basis for a reduction of semantic or ontological notions to syntactic ones: the claims of both involve semantic vocabulary (in particular the notions of truth and truth-preservation). I do not believe that Frege's context principle licenses any kind of "syntactic priority thesis", according to which semantic relations and the corresponding ontological categories are to be reduced in terms of syntactic categories and our responses to syntactically individuated bits of language. This sort of reductive view is adopted by some neo-logicians — see, for example, Wright (1983). It is hard to see how this brand of neo-logicism can respect one of Frege's fundamental aims — namely, to show that logical knowledge is *substantive* a priori knowledge. If truths of logic simply reflect the structure of our own syntactic manipulations, then logical knowledge would seem not to be substantive in Frege's sense.

disagreement with Kant over the fruitfulness of pure general logic. But it remains to be seen how such a priori knowledge is possible. Frege does not say very much directly concerning this question. He does, however, say enough for us to be able to attribute to him a quite interesting epistemology of the a priori.

First of all it is important to make the question at issue more precise. On the one hand, Frege's question — as should be expected — is not about the mechanism through which we recognize a priori truths. As he puts it:

Leaving aside logic, one can say: we are forced to make judgments by our nature and external circumstances, and if we make judgments, we cannot reject this law — of identity, for example [...]. I do not wish to either dispute or endorse this view [...].³⁸

How it comes about that we actually do judge in accordance with the law of identity is of no interest to Frege, because he takes it to be a merely psychological matter. But neither can we take the question at issue here to be merely metaphysical: the question is not what *makes* the laws of logic true — if, indeed, such a question makes sense in the context of Frege's thinking — but what justifies *us* in taking them to be true.³⁹

It is important to see that on certain views of these matters, and in particular on the Russellian view, this question can prove extremely embarrassing. According to Russell laws of logic are, in the most general sense, laws concerning relations among concepts (or “universals”, in Russell's later terminology) and, accordingly, among the propositions that contain them as constituents. But propositions and their constituents have no internal relation to our minds. If this is so, then it seems that the way to acquire knowledge about them would have to be based on a *contingent* relation between them and our minds. But what could that contingent relation be?

³⁸ Frege (1997a: 204).

³⁹ Burge (1992: 312-313) gives a very clear and convincing account of this point. Apart from the passage just quoted in the text, Burge discusses a passage from Frege (1980: §3) where Frege explicitly describes his project of showing arithmetic to be analytic as an investigation into the “ultimate ground upon which rests the justification for holding it to be true”.

Propositions and concepts are, after all, supposed to be non-spatiotemporal, causally inert objects. A Platonism of this sort can only posit such a contingent relation without much hope of explaining its nature any further: thus Russell must simply insist that acquaintance — a relation modeled on our knowledge of our own sensations — also relates us to non-spatiotemporal and causally inert objects.⁴⁰

Fregean thoughts and thought-components are not similarly constitutively independent of minds, however, and thus Frege does not have to appeal to an unexplained quasi-perceptual capacity to recognize a priori truths. As we have seen, the principles of rationality (including the laws of logic, Frege's main concern) are at the same time both the principles of individuation of thoughts and the principles with reference to which what it is to be a thinker is to be understood. It follows that, to the extent that we are able to have knowledge of the structure of our own capacity for thinking, that is a means for acquiring knowledge of logic as well.

More carefully, according to Frege to imagine mental activity that goes on in violation the laws of logic is to imagine a defective form of mindedness — “a kind of madness”, as Frege puts it in a famous passage:

What if beings were even found whose laws of [thinking] directly contradicted our own and therefore frequently led to contrary results even in practice? I would say: here we have a hitherto unknown kind of madness.⁴¹

Thus the justification we have for our logical *practices* — for our conformity to the “laws of thought” — consists simply in their *being* the laws of thought, norms constitutive of what it is to be a thinker

⁴⁰ Russell (1997:86-9). Burge (1992) is motivated by a question of this sort as well.

⁴¹ Frege (1997a: 203). I have substituted “laws of thinking” for “laws of thought” in the translation, since it is clear that Frege here is talking about the process of thinking, not its object. The passage is clearer if we imagine scare quotes around “laws of thinking” in the first sentence, since what Frege is saying in that passage is that this so-called “thinking” would not be thinking at all: it would be some sort of mental activity that fails to connect with thoughts.

This point needs to be handled with some care: even assuming classical logic to be correct, for example, it is not at all obvious that logicians who take themselves to be rejecting some of the principles of classical logic must either suffer from some form of madness, or else fail to actually reject the principles they take themselves to be rejecting. I discuss this in Chapter 2.

at all. But the norms that determine what it is to be a thinker also determine the nature of thoughts, and therefore, by the same token, principles determining what the case is or can be. Thus if thinkers as such have a capacity for recognizing the principles that their thinking is subject to (as Frege seems to suppose they do), then that would also be a capacity for substantive a priori knowledge.

4. This, I think, is about as much of an epistemology of basic logical knowledge as Frege has to offer (he has more to say about knowledge of logic as a whole, of course). The fundamental idea is, I think, quite attractive. The rest of this dissertation will be devoted to exploring approaches to the epistemology of the a priori that follow roughly this idea. In closing this chapter I want to briefly indicate how I plan to proceed.

For one thing, a lot more than what Frege gives us needs to be said to flesh out his fundamental idea into a substantive account. At the core of Frege's account clearly lies the assumption that we can recognize the principles our thinking is subject to, just in virtue of being subject to them. But how exactly is that supposed to work? If the grip the laws of thought have on us is just normative — and Frege gives us no grounds to assert anything other than that — then it is a serious question how we are supposed to find out about them. This question will take center-stage in much of what follows in this dissertation.

Moreover, and unlike Frege, my goal is to explore the prospects of an epistemology of the a priori in general, not just of logic. Frege does not explicitly deal with such matters. Nevertheless, I think the fundamental idea of his epistemology of logic can be extended to areas outside of logic. The leading idea of such an extended Fregean account of a priori knowledge is that thoughts — thus what may be the case — and the capacity to think are internally related to each other. On these grounds, it is plausible that reflecting on one's capacity to think can be a source of substantive knowledge. In the most general case, where reflection only concerns thinking as such (as opposed to thinking this or that more determinate kind of thought), the resulting knowledge will have the

universal applicability of logic. It is not, however, implausible to hold that more determinate kinds of thought will be similarly connected to more determinate kinds of constraint on our thinking. Thus it is not implausible that there can be substantive a priori knowledge, attainable in a way analogous to the one Frege envisages for logic, but which goes beyond logic.

I attempt to sketch a more general account of the individuation of thoughts that would substantiate precisely this sort of generalization of Frege's fundamental idea in an Appendix to this dissertation. From the next chapter, and to the end of this dissertation, I turn to consider explicitly how we should go about constructing an account of a priori knowledge, within this broadly Fregean conception of thinking and thoughts.

Chapter 2 CONCEPT-POSSESSION AND A PRIORI KNOWLEDGE

1. INTRODUCTION

According to what I argued in my first chapter, a Fregean account of content — “thoughts”, in Fregean terminology¹ — makes it possible to see how an account of a priori knowledge could, in principle, be given. On the Fregean account, the individuation of contents is determined by the norms that govern thinking. But since thoughts are what may be the case — those items “for which the question of truth and falsity can arise at all”² — knowledge about their nature is knowledge about the shapes reality can take. Thus knowledge of the norms of thinking can lead to substantive knowledge — knowledge about what is or may be the case, and not just knowledge about how we happen to represent the world in thought or language. This is the core idea of the epistemology of the a priori I defend in this dissertation.

In this chapter I will begin to develop and defend my approach, although a complete picture will not emerge until later. My approach to the a priori falls within the broad category of views Christopher Peacocke has called “moderate rationalism”.³ According to moderate rationalism, the a priori is to be explained in terms of the nature of conceptual thought. This seems like a reasonable starting point — for, according to an alternative, “immoderate” rationalism, our capacity for a priori

¹ In what follows, I will be using the term ‘proposition’ interchangeably with Frege’s ‘thought’, since the former has become the standard in contemporary debates. But it should always be remembered that, on the view proposed in this dissertation, contents are understood in Fregean, rather than Russellian, manner.

² Frege (1997c: 327-328).

³ Peacocke (2000: 255-8; 2004: 152-6).

knowledge would have to be an optional extra, not consequent upon anything already required by the nature of conceptual thought. But, unless we have an independent account of that capacity, what light could be shed by merely invoking it?⁴

It will be helpful to develop my approach by comparing it to that of recent writers on the a priori who also fall within the category of moderate rationalism. In particular, I will argue in §2 of this chapter that many of the attempts to explain the a priori in terms of an account of conceptual thought fail, as they presuppose an account of concept-possession which is problematic and should be abandoned. In §3 I will suggest a different way of thinking about concept-possession. Although this different account of concept-possession can indeed serve as a starting point for an epistemology of the a priori, it is not the whole story; I will close by suggesting what more is needed.

2. THE ANALYTICITY STRATEGY

1. Let us begin by making the issue at hand more precise. Consider, for example, how one could come to know a priori the law of double negation, the classical elimination rule for negation (assuming, for the sake of the argument, that classical logic is correct). Consider, that is, how one could come to know that A follows from $\sim\sim A$ (where A is schematic).

As has been widely noted, an account of this ability that takes as its starting point one's knowledge of some explicit specification of the semantics of negation — such as a truth-table, or a

⁴ As we saw in Chapter 1, Russell held such an immoderate view. More recently, Bonjour (1998) defends an immoderate view as well. Understanding (or conceptual competence) is, according to Bonjour, a necessary condition for 'rational insight' (Bonjour's term for the faculty of a priori knowledge). But it is not at all clear that any amount of rational insight is, conversely, *required* by understanding. Thus it seems that, on Bonjour's view, a conceptually competent subject could nevertheless suffer from blindness of rational insight. Such blindness would, no doubt, be massively debilitating; but, it is not strictly incompatible with conceptual thought. If this is an accurate account of Bonjour's view, then indeed it does not belong in the epistemological tradition discussed in this chapter.

Tarski-style recursive clause — is not satisfying.⁵ For one thing, an epistemology of the a priori presumably would have to explain how one comes to recognize the correctness of the Tarski-style clause or the truth-table, just as much as it would have to explain how one comes to recognize the law of double negation as a valid rule of inference. But, perhaps even more seriously, even if we assume that one knows — for example — the Tarski clause for negation, in order to derive from it the validity of the law of double negation one would need to infer in accordance with the law of double negation itself. But it is clear that *that* inference is not based on one's having derived the law of double negation from one's knowledge of the Tarski clause for negation. It must be based on some capacity more primitive than that.

Something along the lines of the following two-stage picture seems intuitively correct. In the first stage, a subject recognizes that, for particular instances of A , the inference from $\sim\sim A$ to A is a good one. This, I take it, is a manifestation of a distinctive ability to conform one's thinking to the epistemic norms governing the relations between two belief-types: the subject recognizes that a commitment to $\sim\sim A$ would be, other things being equal, a conclusive reason to commit to A .⁶ Now, in the most primitive case, this ability will be manifested in the subject's dispositions to perform, expect, and assent to particular transitions in practice, rather than any explicit beliefs about the validity of any inferences. Once, however, our subject becomes more articulate, and if she is equipped with sufficient competence in the relevant logical concepts, she can recognize that the validity of the inference in any particular case is due to the characteristic contribution of negation to a proposition. But then, she is in a position to know that the validity of the inference depends only

⁵ See Boghossian (2000: 245-254), Peacocke (2003: 117-119).

⁶ It is, perhaps, controversial on my part to put the point in terms of the rational relations between belief-types, rather than relations between propositions. I do not have the space to expand on this choice. But, given the normative formulations that will follow, it is a natural one. To the extent, however, that one thinks that propositions *themselves* stand in normative epistemic relations to each other, which are not simply parasitic on normative relations among beliefs, I think one could reformulate everything I have to say in those terms. (I thank Cian Dorr and Tom Ricketts for urging me to clarify this point.)

on the form of the premise and of the conclusion, rather than the particular choice of \mathcal{A} . Thus, she is in a position to know the law of double negation.

A similar two-stage picture seems correct even in cases where norms of inference are not explicitly in view. Consider, for example, how one could come to know a priori that everything colored is extended (assuming that this is indeed something we can know a priori). A subject in a first stage recognizes that a commitment to an instance of “ x is colored” is, other things being equal, conclusive reason for “ x is extended”. Given enough sophistication and reflection, she can recognize that these inferences are independent of any particular choice of x , and thereby come to recognize them as instances of a valid schema. But, once our subject is in a position to know that the inference from “ x is colored” to “ x is extended” is valid for all x , she is also in a position to know that everything colored is extended.⁷

Given this two-stage picture, the pressing question is to explain the type of knowledge exemplified in the *first* stage of our examples — to explain, for instance, how a subject can recognize that an instance of the law of double negation is correct, even though she is not yet able to identify it as an instance of a valid schema. The idea to explore here is that this knowledge is to be explained

⁷ One feature of this account is that knowledge of a non-modal, non-normative proposition such as “everything colored is extended” may depend on much richer knowledge, namely knowledge of the validity of the inference from “ x is colored” to “ x is extended”. But shouldn’t a sparer piece of knowledge be — in some sense — easier to acquire than a related, but richer one? (I thank Cian Dorr from bringing this question to my attention.)

On the present view, however, there *is* a sense in which the sparer knowledge may be easier to acquire than the richer one, even if the latter is epistemically more basic. That a piece of knowledge is epistemically more basic does not entail that it must be more readily available to us, from a subjective point of view (for example, as a premise in further conscious reasoning). Consider the following analogy. On every occasion of perception, we acquire a mass of non-inferential knowledge, and, based on that, a bunch of inferential knowledge as well. Now, the point to notice is that this inferential knowledge may be much more readily available to us than the non-inferential knowledge it derives from. Suppose I’m attending a soccer match, and I see a goal being scored. My knowledge that a goal has been scored is, no doubt, inferential, and it is partly based on non-inferential perceptual knowledge concerning the visual scene before me. It is clear, however, that a large part of that non-inferential knowledge is nowhere near as readily available to me as my inferential knowledge that a goal has been scored. In the case of a priori knowledge, just as in the case of perceptual knowledge, the epistemically more basic piece of knowledge may sometimes be the less readily available to us.

in terms of the role the law of double negation has in our thinking. Supposing that classical logic is correct, I take it as uncontroversial that our thinking is, in a normative sense, bound by that law. If I am committed to $\sim\sim A$ then rationality requires either that I also be disposed to commit to A , or, if I have compelling reason not to commit to A , that I revise my commitment to $\sim\sim A$. Thus, the pressing question is to give an account of a subject's capacity to recognize a priori particular instances of the norms governing her thinking.

Now, the normative characterization of the role the law of double negation has in our thinking is not, by itself, going to answer that question: being normatively bound by a rule does not even explain a subject's *conforming* to that rule, let alone her doing so because she recognizes the correctness of its instances. What we need to focus on, therefore, is an account of thought that can answer such questions.

2. The approach I will discuss and criticize in this chapter has been proposed in different versions by different authors, among them George Bealer, Paul Boghossian, David Chalmers, Frank Jackson and Christopher Peacocke.⁸ The views of these authors of course differ in significant details; however, they all share a bunch of central commitments which constitute what I will call the “analyticity strategy”. These commitments can be illustrated by considering how they would answer the question about our knowledge of the law of double negation left unanswered above.

Suppose that in order to possess the concept of negation one needs to have certain dispositions to infer in accordance with its classical introduction and elimination rules, including the law of double negation. One way in which these dispositions will manifest themselves, presumably, will be in our subject's judging particular instances of inferences that follow these rules to be correct. Moreover, given sufficient reflection and a general background sophistication, our subject may then

⁸ Bealer (1999); Boghossian (1996, 2000, 2003); Chalmers (1996, 2002a); Jackson (2000); Peacocke (2000, 2003, 2004).

come to recognize that these inferences are correct because of their form, rather than the specific contents of the premises and conclusions. She may, therefore, come to explicitly judge the law of double negation to be a valid rule of inference. Since the subject's judgments in both stages are expressions of her dispositions to infer in accordance with the introduction and elimination rules of negation, and these dispositions are constitutive of possessing the concept of negation, they deserve to count as knowledge.

There is an obvious affinity between this strategy and the idea of analyticity as it has been understood since Kant. According to Kant's definition, a subject-predicate judgment is analytic if the predicate-concept is "already thought in" the subject-concept.⁹ Of course much of the detail of the Kantian formulation is dated, and would not be used today. The main idea, however, seems to be to delineate a class of judgments such that, if one grasps the concepts involved, then one is necessarily led to endorse them. Clearly, it is the same idea that motivates the strategy under question here. We can say, then, that the proposed strategy is to reduce the a priori to the analytic.¹⁰

Although there is a lot that makes the analyticity strategy attractive, I do not think it can ultimately succeed. The problem is with the account of concept-possession it presupposes. To put

⁹ Kant (1998: A6-7/B10-11).

¹⁰ I think the analyticity strategy fails, for reasons I will explain in what follows. This shows only that not *all* of our a priori knowledge is analytic. Of course possessing a concept normally does entail being in a position to know some trivial propositions involving the concept itself. Possessing the concept *chair* entails being in a position to know that it applies to all and only chairs, for example (throughout, I use italics to refer to concepts). Thus our knowledge of this proposition may be analytic. But this is not the kind of a priori knowledge the analyticity strategy hopes to explain. What would be interesting to explain, rather, is the connection between possessing the concept *chair* and knowing some non-trivial specification of what it is to be a chair, or between possessing the concept of negation and knowing its introduction and elimination rules.

Note that the analyticity strategy can be carried out in ways that avoid the familiar Quinean arguments against analyticity, which were in any case directed against a much narrower notion, characteristic of logical empiricism (see Quine (1935; 1954; 1980)). The most convincing of these arguments target a metaphysical distinction between truths made true by the way the world is, and some other truths which are, purportedly, made true simply by convention or stipulation — simply because of the ways in which we decide to think or talk. But there is nothing in the analyticity strategy itself that dictates such a distinction: the strategy concerns the way we come to know certain truths, not what makes them true. It is, in particular, perfectly natural for a proponent of this strategy to insist that what makes a priori truths true is simply that their homophonic truth-conditions obtain (see Boghossian (1996)).

the point in a nutshell, in order for the strategy to work, a certain residual “individualism” or “internalism” about concept-possession is required. The analyticity strategy, of course, does not need the extremely controversial thesis that possessing a concept in general supervenes on states internal to a subject. It does, however, need the weaker thesis that possessing a concept entails a fixed set of dispositions to judge or infer. But, as I will argue, even this weaker thesis is not a compulsory, or even plausible, requirement of concept-possession.

3. We can bring the problem out by considering how proponents of the analyticity strategy would interpret cases of difficult a priori disputes.¹¹ Since, on their view, a priori knowledge is a matter of dispositions necessary for possessing the relevant concepts, it seems that in *every* such case a proponent of the analyticity strategy will face significant pressure to ascribe a failure to grasp the relevant concepts to one or the other party in the dispute. This, however, is in many cases too quick. Ascribing failure to understand the terms of a dispute to one or the other party should be a substantive interpretative decision, motivated — if at all — by an analysis of the details of the case and its historical context. It should not be mandated by a philosophical theory of concept-possession.

We can make all this more precise by considering a specific example. Take a classical logical theorem — for example, an instance of the law of excluded middle: the thought that it is either raining or not raining in Pittsburgh on April 20, 2020. Suppose (for the sake of the argument) that classical logic is correct, and therefore that every instance of the law of excluded middle is true. Now, on this supposition, it is safe to say that if anything remotely interesting is knowable a priori, then any instance of the law of excluded middle is knowable a priori (excluding, if required, instances unthinkable by beings with our limited brain capacity). It can, for example, be proven as a

¹¹ Williamson (2006) also discusses examples of a priori disputes in a similar spirit, although Williamson’s argument differs from mine in significant details.

theorem in a natural deduction system. According to the analyticity strategy, adequately grasping the concepts of disjunction and negation must be sufficient for one to be in a position to know basic logical principles from which to produce or recognize a valid derivation with the law as its conclusion (given sufficient background knowledge, attention, and the like). Contraposing, if one is not in a position to produce or recognize such a derivation, then — other things being equal — one lacks the classical concepts of negation and disjunction.¹²

But this is way too quick. As is well known, there are people who reject the law of excluded middle (and, by implication, some or all of the principles that can be used to derive it) on philosophical grounds. Since knowing a proposition requires believing it, those who reject the law of excluded middle do not know the law of excluded middle. And it seems possible that one may not only reject the law of excluded middle, but also lack any disposition at all to come around; thus, one may fail to be in a position to know the law of excluded middle. According to the analyticity strategy, however, it appears that this can only result from a failure to fully grasp the concepts of negation or disjunction. Yet it is not at all clear that this must always be the case: on the contrary, it seems possible that some of those who reject the law of excluded middle may understand perfectly well what it is they are rejecting: they reject it precisely *because* they understand it.

An intuitionist rejecting the law of excluded middle, for example, would argue in the following way. Any deductive argument for a disjunction must be an argument for one of its two disjuncts: we can have no proof of ' A or B ' that is not a proof of either A or B . But sometimes it is impossible for us to have any deductive argument either for a proposition or for its negation — and any proposition about the weather in a specific place at a specific time years in the future is a case in

¹² One may be tempted to object here that competence in deductive reasoning does not require explicit knowledge of all the principles presupposed by it. In particular, one does not need to have established the soundness of the rules of inference one uses in order to be able to acquire knowledge by their use. This is clearly true, but my argument only requires something much weaker, namely, that such reflective logical knowledge, *when present*, can be completely accounted for by reference to capacities for deductive reasoning entailed by possession of the relevant logical concepts.

point. Since any deductive argument for a disjunction must be an argument for one of its disjuncts, it follows that we can have no argument that either it rains on April 20, 2020 or it does not.

Arguably, the intuitionist's argument, far from resting on a misunderstanding of the logical constants, trades on just their ordinary meanings for whatever power to persuade it may have. On such an interpretation, intuitionists and classical logicians share the same concepts of the logical constants, and their disagreement lies elsewhere — perhaps in the classical logician's insistence that only models that respect bivalence are admissible and the intuitionist's failure to endorse this restriction, for example.

The point here is not to defend this interpretation against alternatives. For all I know, the best way to interpret intuitionism might indeed involve the claim that intuitionists do not use the same concepts of the logical constants as their classical opponents. The point is, simply, that a philosophical theory of content and a priori knowledge should not *force* the issue. Intuitively, such issues should be decided by studying the history of logic and mathematics, not by a philosophical theory of content and a priori knowledge. But the analyticity strategy does not seem to leave room for that.

4. One might protest that there is a certain sense in which the dispute just described is not meteorological. There is, for example, no disagreement over how much water is pouring out of the sky on a certain date. Thus, it may be claimed that the dispute must be over concepts or language, not over mind-independent or language-independent fact. On these grounds, one might attempt to define a notion of understanding such that any dispute that fits this loose characterization will always involve the attribution of defective understanding.

I think there are several things that are suspect in this line of thought; I will not, however, attempt to disentangle it here. I simply want to point out that such a notion of understanding will beg the question in favor of the analyticity strategy, by simply *stipulating* that a priori error is

incompatible with understanding. This, in effect, is how Bealer's notion of "determinate understanding" works. Here is the relevant part of Bealer's definition:

Determinateness = the mode m of understanding such that, necessarily, for all x and all p understood m -ly by x :

(a) p is true *if* it is possible for x to settle with a priori stability that p is true.¹³

To "settle with a priori stability" on a proposition understood m -ly is to believe the proposition in such a way that no improvement of one's cognitive situation or of one's cognitive capacities can shake one's belief without also changing one's mode of understanding.¹⁴

Given the assumption that classical logic is correct, this definition entails that our intuitionist fails to determinately understand the law of excluded middle, since, as we may suppose, the intuitionist has settled upon her mistaken views with a priori stability. But the conclusion to draw from this, I suggest, is simply that the definition is inadmissible in a debate over the validity of the analyticity strategy. What we still lack is an independent argument to the effect that full understanding entails a fixed set of dispositions to judge or infer, not a stipulation that makes such a set a condition of understanding.

5. Given the difficulty of providing such a sweeping argument, a proponent of the analyticity strategy might concede that stable a priori error may be compatible with full understanding, while

¹³ Bealer (1999: 45). I omitted two further conditions Bealer imposes in order to deal with a posteriori property identities and cases of deference. Since our concern here is with logical concepts and subjects who do not defer to anyone else, these conditions are irrelevant. Nevertheless, the kind of objection I raise in the text could be raised against at least the second one of them as well. That condition in effect says that, for every (relevant) truth the subject is capable of believing, she must be able to believe it without thereby switching out of her current mode of understanding. The modalities here are crucial. As I have set the case up, there is a certain sense in which our subjects cannot believe the law of excluded middle. If that's the relevant sense of 'can', then our subjects satisfy this clause (which is as it should be, since their understanding is not deferential). But if the 'can' is supposed to be weaker than that, then this clause stipulates against a priori error, thereby begging the question.

¹⁴ Bealer (1999: 42). The sorts of improvement we are here envisaging involve enhancement of memory, clear-headedness, or acquisition of further relevant concepts. They cannot involve the outright acquisition of further propositional knowledge.

attempting to mitigate the damage by appeal to a notion of implicit knowledge. The notion of implicit knowledge used by philosophers is notoriously slippery, but for our purposes I think the following points capture a serviceable core idea:

- (i) Attributing to a subject S implicit knowledge that p must explain her mental state and her behavior in a way that no weaker state (e.g., mere belief or justified belief) could.
- (ii) If S implicitly knows that p , she need not explicitly believe that p . What this means is roughly captured by the following subjunctive conditional: other things being equal, if asked to consider the question whether p , she might either not understand the question or decline to assent to p .

Now, to return to our example, a proponent of the analyticity strategy might hope to help herself to this notion of implicit knowledge and claim that the intuitionist does, after all, know the classical principles from which the law of excluded middle follows — it is just that she knows them *implicitly*, rather than explicitly. The difference between intuitionists and classical logicians is that, while a classical logician can make explicit her knowledge of the classical principles, an intuitionist cannot: her knowledge remains *merely* implicit, and thus it does not enable her to either give or recognize a proof of the law of excluded middle.

Such an account is clearly suggested by Peacocke (2003). According to Peacocke, our understanding of many concepts, including those of the logical constants, is constituted by implicit knowledge of rules determining their contribution to the truth-conditions of whole thoughts. For example, our understanding of the concept *chair* consists in our implicit knowledge that something is a chair if and only if it is an artifact with a seat and a back that is meant to seat no more than one person, and so on. The implicit knowledge in which a subject's understanding of a certain concept consists is her "implicit conception" of that concept.¹⁵

Implicit conceptions might seem to help with the issues raised by a priori disputes. As long as there is, for example, a clear distinction at the level of implicit conceptions between possessing

¹⁵ Peacocke (2003: 119 and *passim*).

the classical concepts of the logical constants and failing to possess them, the proponents of the analyticity strategy may appeal to it to formulate a substantive interpretative question that is not forced by their account of content and a priori knowledge.¹⁶ Peacocke expresses the point as follows:

Since it can be hard to make explicit the content of one's own implicit conceptions, we should equally not be surprised if thinkers sometimes mischaracterize the content of their implicit conceptions. A thinker's explicit endorsement of an incorrect definition does not mean that he does not have an implicit conception whose content is the correct definition.¹⁷

The intuitionist, as we have been assuming, explicitly adopts deviant rules for some logical constants. According to the present line of thought, however, this does not mean that she misunderstands them: she may have the correct (classical) implicit conceptions, simply failing to make them explicit. The question whether she does possess them is, then, a substantive interpretative question which is not forced by the analyticity strategy.

I believe the appearance that implicit conceptions offer a way out of the problem is an illusion, however. Implicit conceptions are attributed for explanatory purposes. Presumably, we *would* have reason to attribute a classical implicit conception to an intuitionist, *if* she were disposed to judge about particular cases in a classical way, regardless of what general principles she went on to endorse explicitly. But we have no reason to believe that the intuitionist will have classical dispositions to judge about particular cases, and so no reason to attribute to her a classical implicit conception of negation. For one thing, an intuitionist will most likely have arrived at her logical deviance precisely through her intuitive judgments about particular cases: she will have first found it wrong, for example, to judge that either Goldbach's conjecture or its negation is true, and only then

¹⁶ Peacocke's own suggestion of the content of an implicit conception for disjunction is not capable of distinguishing between intuitionism and classical logic (see Peacocke (2003: 119)). However, given Peacocke's account of logical knowledge, it is very pertinent to ask *where*, if not at the level of implicit conceptions, the difference between intuitionists and classical logicians will show up.

¹⁷ Peacocke (2003: 123).

adopted her deviant logic. But even if we consider the case of a subject who came to adopt intuitionism directly through explicit instruction in general principles, it seems wrong to think that her adoption of intuitionism will not have affected her dispositions to make particular judgments. But if it has affected them, then again we lack any reason to attribute to her *now* a classical implicit conception of negation.

In the absence of any disposition to judge classically, I can see only one way to maintain that the intuitionist still has a classical implicit conception of negation. This is to say that the classical implicit conception is the content of a subpersonal state of her reasoning system, whose effect on her personal-level mental activity is for various reasons systematically overridden. This, of course, does not answer the request for *positive* reasons to attribute classical implicit conceptions to a subject who is not disposed to judge classically. But we can waive this objection for a moment, because I think there is another important point to be made in connection with this response.

If this is what having an implicit conception of negation comes down to, then the implicit conception cannot be a source of a priori knowledge, even in cases where its effect on the subject's personal-level mental activity is *not* overridden. A subject's epistemic access to the workings of her own subpersonal systems is epistemically no different from the access another subject — perhaps a cognitive psychologist who uses her as an experimental subject — has to them. Since the latter is evidently a posteriori, so is the former.

Consider what the case would look like from the perspective of the subject. The subject would find herself inclined to respond to certain kinds of input in ways that follow certain patterns, but she would have no access to these patterns except through inductive reasoning on the data-set of her own responses. Since the implicit conception operates only at a level to which she has no direct access, she would have no direct access to what prompts any one of her responses, nor could she see any reason behind them. Whatever rationality she might find to her responses will be

retrospective, based on the patterns she inductively recognizes in them. This, however, is no way to acquire a priori knowledge; it is, at most, a way to acquire empirical knowledge of one's habits of thinking.

This point is perhaps even more relevant in cases of concepts such as *chair* which, unlike logical concepts, are frequently associated with capacities to perceptually recognize objects as belonging to a certain kind. On certain views on these matters, a subject who possesses such a concept in virtue of having the corresponding recognitional capacity may gain a priori knowledge of the corresponding kind by reflection on her recognitional capacity. In the terms employed in this section, the idea would be roughly as follows. A subject has an implicit conception of the concept *chair*, for instance. This conception guides her application of the concept in particular cases. By reflecting on these particular dispositions to judge, then, the subject comes to know a priori that chairs are artifacts with seats and backs, and so on.

I do not want to deny that this might be a way to figure out what it is to be a chair. Still, we need to be careful here, for although it is obviously true that a recognitional capacity must operate on the basis of stored information about the relevant kind, this information may be available only to the subject's subpersonal perceptual systems, and not to the subject herself. The phenomenology of perceptual recognition seems to support this idea: recognizing an object as belonging to a certain kind is coming to know it belongs to that kind, even in the absence both of an account of what it is to belong to that kind and of any reasonably detailed mental images of exemplars of that kind. If this is the case, then the subject is not able to derive any a priori knowledge from her subpersonal implicit conception. Perhaps she may still be able to derive some kind of knowledge about the relevant kind, by inductive reasoning on the data-set of her own dispositions to classify particular objects as belonging to it. But this is not a priori knowledge of what it is to belong to that kind. In

effect, the subject is simply using herself as an instrument for the detection of instances of the kind in her environment.

We can put the problem with the appeal to implicit conceptions in the form of a dilemma. Implicit conceptions are either wholly subpersonal, or not. But they cannot be wholly subpersonal if they are to play any role in the epistemology of a priori knowledge. If they are not, however, then it is extremely implausible that an intuitionist will have classical implicit conceptions of any of the logical concepts that play a role in the disputed principles. Therefore, it appears that the analyticity strategy still entails that the intuitionist and the classical logician cannot share the same logical concepts. If, as I have been suggesting, it is bad interpretational practice to prejudge this issue on the basis of abstract philosophical theory, this gives us reason to doubt the analyticity strategy.

6. Throughout, I have been assuming that the dispute between classical and intuitionist logicians is a priori — just as the participants in the dispute take it to be. It may be suggested, however, that the (seeming) intractability of the dispute, together with the lack of independent arguments that some of the participants must lack the relevant concepts, provides us with enough evidence to conclude that the dispute is not, after all, a priori. If this were right, then the debate between intuitionists and classical logicians would no longer pose a threat to the analyticity strategy: it would now be possible to claim that both intuitionists and classical logicians possess the same dispositions to judge and infer, which, however, are insufficient to resolve the dispute.¹⁸

¹⁸ The possibility of this move was pointed out to me by John McDowell. This move would, in effect, constitute a vindication of the intuitionist position. The intuitionist claims that we have no right to assign a truth value to an instance of the law of excluded middle, unless we know the truth-value of its disjuncts; and in many cases (such as our earlier example about the weather on a particular day in Pittsburgh), this will only be possible on the basis of empirical evidence. It may therefore be suggested that sound interpretational practice tells against this suggestion, since it tells against forcing a resolution on an apparently substantive debate in such a high-handed manner. I think this is a good point; yet given the problems with the previous attempts to defend the analyticity strategy, the present suggestion deserves to be considered in more detail.

I think the present suggestion is problematic, because it introduces an *ad hoc* interpretational principle that creates many more difficulties than it solves. The principle that seemingly intractable disputes cannot be a priori is far too sweeping. This is of course not to say that we may never in practice take the appearance of intractability as one kind of evidence that a certain dispute may not, in fact, be a priori. The objection is only against the claim that such considerations are sufficient to show a dispute not to be a priori.

We can begin to see the problems by returning to the example of the law of excluded middle. Consider an instance of the law of excluded middle ' $p \vee \sim p$ ', where p is empirical, and the question whether this particular instance can be known a priori or not. According to our principle it cannot be known a priori, since it is a thought disputed by subjects who suffer from no relevant misunderstanding. Nevertheless — again according to the same principle — it cannot be known a priori that it is not knowable a priori, for the classical logician believes that it *is* knowable a priori, without misunderstanding any of the relevant concepts.

More precisely, the classical logician believes the following (for some empirical p):

(KA) It is knowable a priori that $(p \vee \sim p)$.

According to the current proposal (KA) is false, and so of course not knowable by any means. Yet the classical logician need not be guilty of any misunderstanding: we have assumed she does not misunderstand the logical concepts, and we can also stipulate that she misunderstands none of the epistemic concepts involved either. The intuitionist, on the other hand, disputes (KA), claiming that it is false (name this $(\sim KA)$). We can again stipulate that the intuitionist does not misunderstand any relevant concepts. Now, the true thought $(\sim KA)$ is disputed by subjects none of whom misunderstand any of the relevant concepts. Thus, again according to the current proposal, it is not knowable a priori.

The conclusion generalizes to every case that shares the same structure. This by itself should be uncomfortable for one who chooses to defend the analyticity strategy in this way. Even worse, however, it seems that the proposal leads to incoherence. For it seems that we do not need *actual* disputants in order to generate the conflict that forces the conclusion that the thought at issue is not knowable a priori. Imagine the epistemic situation of a subject who has never had any empirical evidence for the existence of people who reject the law of excluded middle. It seems possible for such a subject to go through the relevant reasoning in her armchair, thereby concluding that (KA) is false. Such a subject may imagine a character producing the intuitionist arguments against the law of excluded middle, and rationally conclude that such a character need not misunderstand any logical concepts. Supposing that she herself believes in classical logic, and that she does not feel persuaded by the imagined intuitionist's arguments, she would reasonably conclude that the law of excluded middle is not knowable a priori. Would the proponent of the present strategy then say that she *knows* that the relevant instance of the law of excluded middle is not knowable a priori? This cannot be right, because then she would know $(\sim KA)$ *a priori*: her reasoning nowhere relied on empirical evidence. As we have seen, however, $(\sim KA)$ is not knowable a priori. Yet what independent grounds do we have to deny she knows $(\sim KA)$ just in virtue of her reasoning?

Even if there is a way to avoid this apparent incoherence, it seems clear that adopting the present interpretation gets us into difficulties we could avoid, if only we took the disputed thoughts to be knowable a priori — just as all the parties in the dispute have taken them to be all along. The present interpretation has nothing to recommend it other than that it helps shield the analyticity strategy from a certain objection. It should, therefore, be rejected as *ad hoc*.

3. A HOLISTIC CONSTRAINT ON CONCEPT-POSSESSION

1. I have so far argued that the analyticity strategy fails because it presupposes an implausible view of concept-possession. According to that view, possessing a concept is a matter of having a fixed set of dispositions to judge or infer in accordance with certain norms. My argument was that principles of sound interpretation tell against this view: whether or not a subject possesses a given concept should be decided on the basis of an interpretation of her total cognitive state that makes best sense of her interactions with her physical and social environment, rather than by looking at a narrow segment of her psychological dispositions. If we reject this premise of the analyticity strategy, there is no difficulty in accepting that whether deviant logicians share the same concepts as their classical opponents is a substantive interpretative question, not to be settled by a philosophical theory of content and a priori knowledge. This is so even if our deviant logicians do not have classical dispositions to judge or infer, and therefore fail to be in a position to know some of the classical principles governing these connectives.

None of the above points is inconsistent with an explicitly normative thesis, according to which to possess a concept is to be subject — *normatively* subject — to the norms that fix its place in our rational economy, and thereby serve to individuate it. In the previous chapter I suggested an account of Fregean thoughts that directly links the individuation of contents to the norms of rationality that govern the thinking of both individual thinkers and intellectual practices. On such a view, the normative thesis that to possess a concept is to be subject to norms that serve to individuate it follows directly from the nature of content.¹⁹

¹⁹ It is important to note, however, that the normative thesis about concept possession can be motivated independently of the details of the view of content adopted here. In general, as long as contents are conceived of as ways of carving up logical space (to use a familiar metaphor), they will be individuated in terms of their relations to each other. To the extent that one takes contents to be structured in the appropriate sense, these relations will be further resolved into relations among thought-components (or concepts). Moreover, it is natural to think that these relations among contents or among concepts will

The crucial question, of course, is what exactly it takes for the thinking of a subject to be subject to the norms constitutive of a particular concept: the normative characterization of concept-possession does not by itself take us any closer to explaining the possibility of a priori knowledge. In broad terms, what it takes for a subject to possess a concept is for her to participate in a corresponding norm-governed practice. In the next section I will make some general remarks about what participating in such a practice is; in the section after that I will suggest that this approach can help ground an epistemology of a priori knowledge. I will close by elaborating on some aspects of this idea, although the full account will have to wait for a later chapter.

2. Our starting point is that one can participate in a practice which, as a matter of normative fact, is bound (for example) by the classical introduction and elimination rules of negation, and therefore one can count as possessing the concept of negation, while lacking the disposition to conform to all of these rules. This is not to deny that there are substantive conditions necessary for membership in a norm-governed practice; it is simply to deny the premise of the analyticity strategy, according to which possessing a concept entails a fixed set of dispositions to judge or infer in accordance with its constitutive norms.

The spirit of a view of concept-possession along the lines that will be proposed here is captured by Timothy Williamson, in the following very suggestive passage:

What binds together uses of a word by different agents or at different times into a common practice of using that word with a given meaning? This is an instance of a more general type of question: what binds together different events into the history of a single complex object, whether it be a stone, a tree, a table, a person, a society, a tradition, or a word? In brief, what makes a unity out of diversity? Rarely is the

correspond to normative relations among propositional attitudes. (This is part of the point of Jackson (2000). Jackson somewhat spoils this good point by moving directly from the normative thesis to a thesis about what a subject can know in virtue of possessing the corresponding concepts.) The logical relations between A and $\sim\sim A$, for example, correspond to the normative relations between beliefs with A and $\sim\sim A$ as their contents. The normative thesis about concept-possession above amounts just to the claim that possessing a concept entails being capable of propositional attitudes that stand to each other in the appropriate normative relations.

answer to such questions the mutual similarity of the constituents. Almost never is it some invariant feature, shared by all the constituents and somehow prior to the complex whole itself [...]. Rather, it is the complex interrelations of the constituents, above all, their causal interrelations. [...] The claim that a shared understanding of a word requires a shared stock of platitudes depends on the assumption that uses of a word by different agents or at different times can be bound together into a common practice of using that word with a given meaning only by an invariant core of beliefs. But that assumption amounts to one of the crudest and least plausible answers to the question of what makes a unity out of diversity. In effect, it assumes that what animates a word is a soul of doctrine.²⁰

What makes it the case that a group of subjects express the same thoughts by the same form of words, in spite of possible disagreements — and even difficult a priori disputes between them — is that they all participate in a common practice of using these words with a common meaning, a practice which itself is constituted by the different, and often asymmetric, contributions all of its participants make.

Williamson's passage concerns specifically words in a natural language, but it can be extended to the case of thoughts and concepts as well. Our disagreeing logicians may, for all we have seen, all share the same concept of negation and disjunction. The explanation, if we assume this to be true, would not be that they share a specific set of abilities or dispositions: it would, rather, be that they are all members of the same practice, drawing from its "shared stock of thoughts", as Frege would say.

So what is the relation between the practice, conceived of as constituted by the totality of the dispositions of, and the interactions between, its individual members, and the norms constitutive of the stock of thoughts available to its members? The case of our disagreeing logicians was rather exceptional in this respect, for it was assumed that it contained some members who actually *had* knowledge of the relevant norms: the introduction and elimination rules endorsed by classical logicians were assumed to reflect the true nature of the concepts involved. This, however, need not always be the case. It is possible for a practice to be subject to certain conceptual norms, even

²⁰ Williamson (2006: 45).

though no members of the practice are able, on the basis of conceptual and theoretical materials readily available, to articulate these norms. Peacocke discusses two alleged such examples, drawn from the history of mathematics (whether the examples are historically accurate or not is less relevant to the present point than the fact that they seem to illustrate genuine possibilities).

The mathematical concept of limit was explicitly defined in an adequate way only by Weierstrass in the 19th century; yet since the time of Newton and Leibniz, people who had no knowledge of this definition had been able to calculate the limits of specific sequences or functions in a more or less stable fashion. Arguably, according to Peacocke's view, some or all of these people were using the same concept, and were therefore committed to the same norms, even though none of them was in a position to articulate the content of these norms.²¹

The concept *whole number* arguably provides us with another example. Peacocke suggests that our number-thoughts are constituted by the following set of rules:

- (1) 'is a whole number' is true of 0;
- (2) 'is a whole number' is true of the successor of anything it is true of; and
- (3) nothing falls under 'is a whole number' unless it can be shown to do so on the basis of rules (1) and (2).²²

Yet these rules could not have been formulated without taking into account quite recent mathematical developments. The rationale for the limiting clause in rule (3), for example, would be unintelligible independently of the discovery that first-order axiomatic formulations of arithmetic can be satisfied by non-standard models, whose domain includes objects other than the natural numbers. Nevertheless, present-day formulations of the rules governing the concept *whole number*

²¹ Peacocke (2003: 121-2). Peacocke actually only claims that Newton and Leibniz themselves possessed the concept *limit*; but evidently the practice they initiated went on independently of them. Of course, in Peacocke's view, those who shared the concept of a limit actually *did* know the correct definition, *implicitly*.

²² Peacocke (2003: 126).

aim to capture norms arithmetical thought was subject to, even before their articulation was within reach.

According to the view proposed here, what would settle it that mathematicians throughout have been using the same concept of a limit or of a whole number need not be their sharing the same dispositions to judge or infer, or their possessing the same implicit knowledge of the correct definitions. In fact, the claim that early mathematicians possessed these concepts need not entail, on the present view, that we can non-circularly specify anything very specific that they all shared. The claim — assuming it is well-founded — may be based only on the historical continuity of the practice of mathematics.

In the discussion so far, I have been taking it for granted that there is always a determinate answer to the question of which concepts are operative in a practice. But I do not think there is any reason to rule out from the outset the possibility of indeterminacy here. We have, for example, by now discovered various non-equivalent concepts of limit. Must it be possible to give a determinate answer to the question which one of these concepts was shaping the practice of mathematics and physics before such matters even started to become clear? I do not think it obvious that the answer has to be ‘yes’ in every such case. We are all accustomed to the idea that our thinking about the world is infected by vagueness; there is equally no reason to expect perfect sharpness of our thinking about thought.²³

It would be a mistake to conclude from this, however, that we can never give a determinate answer as to what conceptual norms a practice is subject to, unless some of its members are able to do so too. I think it is intuitive enough that a practice may be subject to determinate conceptual norms which none of the practitioners is in a position to articulate — at least so long as there are

²³ Alternatively, one may suggest that early mathematicians determinately used one concept, which however was itself vague (and thus is not to be identified with any one of the modern concepts of limit). I have no reason to dispute that cases of this sort may arise as well. Exploring such questions at the length they deserve simply falls outside the scope of this dissertation.

other aspects of the practice that distinguish between distinct candidates. The case of thinking about whole numbers seems likely to fit this description. Even though the discovery of non-standard models of first-order arithmetic is recent, and rules such as Peacocke's could not have been formulated earlier than that, a determinate affirmative answer as to whether the non-standard models were ruled out by the practice of earlier mathematicians is plausible — this is why, after all, they were called 'non-standard' once discovered.

It may be helpful to situate my view in relation to Tyler Burge's interpretation of certain relevant aspects of Frege's notion of sense.²⁴ According to Burge, Frege took sense not to supervene even on the totality of the dispositions and abilities of the entire community, specified non-intentionally but with the social and physical environment in view. This is because — according to Burge's account — which Fregean sense is attached to an expression in use must always be fully determinate, even in cases where there is nothing about the dispositions of the members of the community to distinguish between distinct conceptual alternatives. On this view, even if there were nothing at all in the practice of mathematicians and physicists prior to the rigorization of analysis to distinguish between the modern concepts of limit, there would still be a determinate answer as to which one of these concepts was operative in that practice (of course, *we* might still not be able to know that answer).²⁵

This is a strong view, and I am not sure it should be endorsed. I think it is plausible that sometimes what appears to be merely an articulation of a concept already in use actually constitutes a genuine conceptual innovation, or at least a choice between genuinely distinct conceptual alternatives. The history of mathematical thinking about limits may provide an example of this. If we allow for the possibility of indeterminacy about which sense is associated with an expression in

²⁴ Burge (1990). I will not here consider the question whether Burge is right to attribute this view to Frege.

²⁵ On Frege's view, of course, there are no such things as vague concepts, so the alternative envisaged in n. 23 above is not in play at all.

use we can make sense of such cases. Moreover, if we reject the view Burge attributes to Frege, there is no obvious motivation to deny that sense supervenes on the totality of the abilities and dispositions of the community (even if it does not supervene on what any members of the community are in a position to articulate).

3. The questions about vagueness or indeterminacy that arise are important, but I am not going to consider them in this dissertation. The rest of this chapter will focus on the possibility of constructing an account of a priori knowledge on the basis of a view of conceptual practices along the lines just sketched.

This view, as we saw, rejects the premise of the analyticity strategy, according to which possessing a concept entails a fixed set of (non-trivial) dispositions to judge or infer. If we reject this premise, however, it may appear that there is no prospect of taking the nature of conceptual thought as the basis of an account of a priori knowledge. This may be the conclusion Williamson draws from his employment of arguments from logical deviance, in some respects similar to the ones employed earlier in this chapter:

The strategy [...] was to reduce a question in the theory of knowledge to questions in the theory of thought and meaning. If the attempt fails, as it apparently does, then this is more evidence for the autonomy of the theory of knowledge with respect to the theory of content.²⁶

Williamson does not explain exactly what he means by ‘reduction’ and ‘autonomy’. It may, therefore, be unfair to take him as a representative of the line of thought I am about to reject. In any case, it is not hard to imagine someone reasoning roughly as follows:

We are looking for an account of a priori knowledge in terms of the nature of thinking. Presumably, part of this account will be to show that it is no accident that, if one is a thinker, one is able to acquire the relevant a priori knowledge. It follows that it must be shown that, *necessarily*, if one satisfies the conditions laid out by our account of thinking, one is in a position to acquire the relevant a priori knowledge. Yet the argument of the earlier sections, and the conclusion just drawn, show that

²⁶ Williamson (2003: 47).

there are no necessary entailments from concept-possession to being in a position to acquire non-trivial a priori knowledge. It follows, therefore, that no explanation of a priori knowledge in terms of the nature of thinking is possible.

I accept the claim that in order to explain a priori knowledge in terms of the nature of thinking one must show that the ability to acquire a priori knowledge is not an accident, given the corresponding account of thinking. It is also true that, given the rejection of the analyticity strategy and the underlying account of concept-possession, being in a position to acquire a particular bit of a priori knowledge is not plausibly entailed by possessing any particular concept. What I do not accept is the inference from the claim that the ability to acquire a priori knowledge is grounded in the nature of thinking to the requirement of necessary entailments between concept-possession and being in a position to acquire a priori knowledge. I think that an epistemology of the a priori can be based on the nature of thought even without such entailments.

In fact, this claim is just an instance of a more general thesis: an explanation in terms of the essence or nature of a kind or a capacity does not need to consist in the derivation of necessary truths about all members of the kind, or all exercises of the capacity. Such truths are not generally available, and, moreover, insistence on them often misrepresents the sort of explanation at issue. Explanations in terms of the nature of a kind or a capacity generally do not work by ruling out outcomes incompatible with the *explanandum*, but by demoting them to the status of defective, or abnormal, specimens.²⁷ The truth of this general thesis, of course, is not the issue here. In the rest of this chapter I give an outline of the positive view I suggest.

As we saw, the analyticity strategy appears to go wrong in conceiving of concept-possession as a matter of possessing a fixed set of dispositions to judge or infer. On the view proposed here, possessing a concept is a matter of participating in a norm-governed practice. Rejecting the premise of the analyticity strategy entails that membership in such a practice does not require a fixed set of

²⁷ See Thompson (1995) for more on this kind of explanation.

dispositions to judge or infer in accordance with its norms. But it does not entail that membership in such a practice comes cheaply: as I suggest, it is still required that a member of the practice be sufficiently reliable, *on the whole*, in conforming to enough of its norms. The point to notice here is that this constraint is holistic, in the sense that it does not entail that the subject must possess any particular fixed set of dispositions to judge or infer. It is, therefore, immune to the arguments against the analyticity strategy considered earlier.

This holistic constraint follows naturally from taking concept-possession to be a matter of participating in a norm-governed practice. In attributing a particular content to a particular subject we must interpret her total mental state in such a way that it naturally embeds in the corresponding conceptual practice. If the subject's system of thinking showed no intelligible tendency to conform to the norms constitutive of a practice, then we could not make sense of the suggestion that she nevertheless is a participant in that practice. In fact, the harder it is to interpret the subject as conforming to the norms of the practice, the more tenuous her claim to be a participant in it becomes. Moreover, it is important to note that the constraint applies to how a subject is disposed to judge in future or counterfactual cases as well, and not only to her actual performances. In order for a subject to be a participant in a norm-governed practice not only must she have an adequate actual track record, she must also be sufficiently disposed to go on conforming to the norms of the practice in a reasonably reliable way. This, of course, does not mean that she cannot be misled on some possible occasions. Still she must, for a sufficiently broad range of more or less normal conditions, be disposed to go on in the right way.²⁸

²⁸ Using the language of possible worlds, we need a notion of "normal conditions" to pick out a set of possible worlds in which our subject must have an adequate track record in order to count as a participant in the practice in the actual world. Although this notion of normality is not mere statistical normality, it is nevertheless quite intuitive, and in any case it seems necessary for a modal account of pretty much any dispositional property.

The constraint is holistic, in the sense that there need be no privileged subset of the norms of the practice that the subject has to get right in order to count as competent with some particular concept or other. Errors in some sophisticated, difficult logical questions will not count much against our subject's claim to participate in the practice, if she has a good track record in the more basic, simpler cases. But mistakes even in central and basic norms of the practice may also be offset, if they can be explained in a larger context. On this account, therefore, we cannot decide whether our disputing logicians share the same concepts or not independently of a close study of the debate between intuitionists and classical logicians, in its historical context. Even assuming that classical logic is correct, and that the deviant logicians genuinely lack some of the classical logicians' dispositions to judge or infer, it still does not follow that their deviance must be attributed to their lacking the classical concepts. The question, on the present view, will need to be decided by considering which of the two competing interpretations makes better sense of the debate. I cannot propose a definitive list of rules for resolving such questions here, but neither do I think that the lack of such a list renders my claim empty or obscure. Any account of interpretation must, in any case, face these questions in one guise or another.

The crucial point for the purposes of this dissertation is that adopting this weaker, holistic account does not deprive us of a route to explaining how a subject may be in a position to have some a priori knowledge. Consider the case of a subject who is in a position to know a priori a basic logical law. There is nothing, in principle, to stop us from using the subject's disposition to recognize particular instances of the introduction and elimination rules of the constants involved in that law to explain her knowledge of the law. (Since the question is to explain the subject's *knowledge* of the law — and not merely her capacity to understand it — we can assume that she has the appropriate dispositions.) Suppose that the classical introduction and elimination rules are indeed constitutive of our practice with negation and disjunction. Then, a subject's reliable disposition to

recognize inferences in accordance with those rules as correct can — *if* present — still ground her more sophisticated capacity to recognize the rules themselves as valid, and even to employ them in a proof of the law of excluded middle. The fact that some other subject possesses the same concepts of negation and disjunction while *lacking* the appropriate dispositions (and therefore while unable to give or to recognize such a proof), does not disturb our subject's claim to the proof.

4. Now, admittedly, the holistic constraint on concept-possession proposed here does not on its own suffice for giving a full account of our subject's a priori knowledge. The reason why it does not, however, has nothing to do with the analyticity strategy. Although this is a point I will expand in Chapter 4, I will note here that the reason is simply that being reliable in conforming to a certain norm is not yet being able to recognize particular instances of the norm; and it is the *latter* ability, rather than the former, that is the basis of our a priori knowledge. Even though, as we have seen, in the most primitive case one's recognition of particular instances of the norms manifests itself only in dispositions to infer and to judge (as well as to assent to others' judgments and inferences) in the appropriate ways, it is not the case that *every* disposition to make transitions or form beliefs that conform to the relevant patterns manifests any kind of a priori knowledge. A subject may be reliably disposed to make transitions in accordance with the law of double negation, for example, but not because she recognizes such transitions as good ones: the disposition may be the result of a quirk in the wiring of her brain, and the subject herself may be totally in the dark as to *why* she is disposed to make transitions that follow these particular patterns. We can even imagine a scenario in which she is completely mistaken about the patterns her transitions actually display. But such "blind" dispositions are not sufficient to ground any kind of knowledge — or so I suggest.

As long as we lack an account of what it is for a disposition to judge or infer *not* to be blind in this way, therefore, we lack an account of a priori knowledge as well. Accordingly, the final step in my account of a priori knowledge is an account of self-knowledge. A discussion of self-

knowledge, together with a fuller discussion of its role in an account of the a priori, will have to wait for Chapter 4. In the next chapter I turn to consider a newly prominent version of the analyticity strategy — namely, the two-dimensionalist view proposed by David Chalmers and Frank Jackson.

Chapter 3
TWO-DIMENSIONALISM AND THE EPISTEMOLOGY OF RECOGNITION

1. INTRODUCTION

One version of the analyticity strategy that deserves special discussion — both because of its prominence in contemporary debate and because of the picture of the mind it embodies — is to be found in the writings of Frank Jackson and — especially — David Chalmers on two-dimensionalism. Both of these authors explicitly defend an account of conceptual thinking according to which possessing a concept is a matter of a subject's being in a position to know through reflection alone a set of conditional propositions, which concern the concept's application depending on what the world turns out to be like.¹ Such knowledge is supposed to be a priori precisely because it depends only on the subject's possession of the relevant concepts.

Given its reliance on the idea that concept-possession entails knowledge of a fixed set of propositions, this account is of course subject to the same sorts of criticism that — according to what I argued in Chapter 2 — other versions of the analyticity strategy face as well. In the present chapter I will not repeat the same arguments, but rather focus on the picture of concept *application* implicit in the two-dimensionalist account. Concept-application in experience depends on the exercise of a variety of capacities, many of which can be grouped together under the general label “recognitional”: we recognize particulars as instantiating kinds, for instance, or as being identical to

¹ The formulation used here is not accidental, but rather explicitly defended in Chalmers (2002a). I will say more about this later.

things we remember from past experience. But, as I will argue, two-dimensionalism is committed to an implausible account of recognition.²

2. THE TWO-DIMENSIONALIST PICTURE OF CONCEPTUAL THOUGHT

1. The two-dimensionalist account is supposed to apply, with appropriate modifications, both at the level of thought and at the level of language. I will only focus on the case of thought here.³

According to two-dimensionalism, to possess a concept C is to know, or to be in a position to know by reflection alone, what I will call its set of application-conditionals — that is, a set of material conditionals of the following form:

$$D \rightarrow E(C)$$

Here D is supposed to be a variable ranging over exhaustive descriptions of what — given all we know independently of experience — the actual world may turn out to be like, and $E(C)$ a specification of C 's extension, on the hypothesis that D is true of the actual world. Given the set of world-descriptions \mathcal{W} , to every concept corresponds a set of application-conditionals, each of which has as its antecedent a member of \mathcal{W} and as its consequent a specification of what, if anything, the concept applies to under the supposition that the actual world turns out to conform to the description in the antecedent.

² Much work on two-dimensionalism has been published recently. What follows has points of contact with some of it, particularly with Yablo (2002). I will elaborate on this later on in the chapter.

³ A note about terminology is in order here. According to Chalmers's usage, the term 'concept' applies to mental representations, that is, mental particulars with semantic properties analogous to those of words (see Chalmers (2002c: 3)). This is somewhat surprising, since on Chalmers's Fregean view it would be more natural to take concepts to be non-mental, abstract entities. What could be considered as mental is the state of grasping a concept; but then, that is most naturally construed as general rather than particular, since it is a state many different subjects can be in at many different times. In what follows I will use the term 'concept' to refer not to mental particulars, but rather to components of thoughts that can be shared by many subjects.

Continuing in the same spirit, we can define for every concept its “epistemic” (or “primary”) intension.⁴ The epistemic intension of a concept C is a partially defined function from \mathcal{W} , such that for any D — a member of \mathcal{W} — on which it is defined, its value is the extension determined by C ’s application-conditional with D as its antecedent. The epistemic intension of a concept should be distinguished from its “counterfactual” or “secondary” intension. The counterfactual intension is defined in the way familiar from ordinary (one-dimensional) modal semantics, and determines what the concept applies to in counterfactual situations. Thus, to take a standard example, the counterfactual or secondary intension of *water* is a constant function, whose value is the same stuff — H_2O — in every world where it is defined (throughout, I use italics to refer to concepts). By contrast, since (for example) we cannot a priori exclude the possibility that it may turn out that there is actually no such stuff as H_2O , the value of the epistemic or primary intension of *water* is not always H_2O . There are ways the world (the actual world) may turn out to be on which the stuff the concept *water* applies to is not H_2O .

Grasping a concept, knowing or being in a position to know its application-conditionals, and grasping its epistemic intension are supposed to be the same state, according to the two-dimensionalist picture.

Of course no actual subject is capable of entertaining anything like the exhaustive description of the actual world that D purports to be, let alone evaluating any conditional with D as its antecedent. The two-dimensionalist’s claim, therefore, must be not that a subject possessing C actually *knows* any such conditionals (since knowledge intuitively entails understanding), but rather something like the following counterfactual: just in virtue of possessing a concept C , a subject

⁴ My identification of primary and epistemic intensions here follows Chalmers (2006). In earlier work Chalmers took the identity to be something of an open question, to be established by argument (see Chalmers (2002a, 166-7)).

possesses all that would be required in order for her to know the corresponding application-conditionals, if it were not for limitations of memory, computational speed, life-span and the like.

Now, if *D* is allowed to involve any concepts we like, then the two-dimensionalist's claim will be rather uncontroversial. This is because in the absence of any restrictions on the concepts we are allowed to draw upon, we can always employ *C* itself; and in this case, the application-conditionals will be no more informative than a homophonic specification of *C*'s satisfaction-conditions. But it is relatively uncontroversial that a subject who is minimally reflective, and possesses the intentional concepts required to formulate homophonic satisfaction-conditions, will normally be in a position to know such conditions for any concept she possesses. If the two-dimensionalist claim is to be interesting, therefore, restrictions must be imposed on the concepts *D* is allowed to involve. At a minimum, we must not be allowed to use *C* itself, or concepts too closely related to it, in the formulation of *D*. Chalmers imposes much tighter restrictions, claiming that the vocabulary of fundamental physics (or some future refinement thereof), together with purely phenomenal concepts, suffice for the formulation of application-conditionals for all of our concepts.⁵ This is obviously an extremely controversial thesis, and I will not be concerned with it in this chapter. I will remain agnostic as to what exactly is allowed in *D*, requiring only that application-conditionals must be more informative than the corresponding homophonic satisfaction-conditions.⁶

For present purposes, at least, what is most interesting about this account of concept-possession are its implications for the epistemology of a priori knowledge. As we already saw, according to this picture possessing a concept *C* entails being in possession of everything required to

⁵ See, for example, Chalmers (2001). Of course two-dimensionalism must offer a different account of what it is to possess the concepts included in the basic vocabulary — perhaps by allowing homophonic application-conditionals for these concepts.

⁶ Another requirement is that application-conditionals must use only “semantically neutral” vocabulary — they must not, in other words, employ any terms whose content depends on which world is treated as actual (see, for example, Chalmers (2002a: 166), Chalmers (2002b: 17, fn 5)).

know its application-conditionals by reflection alone. Of course *we* cannot know these conditionals, since we are not in a position to entertain the exhaustive world-descriptions in their antecedents. Nevertheless, it seems very plausible that this capacity can put *us* in a position to know a priori a lot of more restricted, but still non-trivial propositions: we can, for example, presumably know a priori — just on the basis of our possession of the concept *water* — that the clear, potable stuff that fills lakes and rivers is water.

2. It is clear that two-dimensionalism is committed to a kind of individualism (or internalism) about conceptual thought. Possessing a concept, according to two-dimensionalism, is a matter of capacities which are constitutively independent of one's relations to a physical or social environment. This, as I will suggest, is what is wrong with the view.

Two-dimensionalism of course concedes something to externalism, by granting that the counterfactual intension of many of our concepts is constant, with its value fixed by the subject's relations to her environment (since it is, as we saw, fixed by the concept's extension in the actual world). This entails that conditions specifiable at the level of an individual subject will not, in general, be sufficient for concept-possession. A subject who has lived all her life on Putnam's Twin Earth (thus lacking all direct or indirect causal contact with H₂O), for example, cannot express the same concept by her word 'water' as we do, even though all her individualistically specifiable dispositions may match those of a subject on Earth.⁷

This concession, however, does not substantially alter the individualistic character of the account. We should remember that possessing a concept is not a matter that directly involves its counterfactual intension — it is a matter of knowing, or being in a position to know, its application conditionals. Of course, to the extent that a concept is *used* — as opposed to merely mentioned — in its application-conditionals, a subject's being in a position to know a specific set of conditionals

⁷ See Putnam (1975).

implicitly settles which world-description in W is true of the world that is actual for her — it settles, in other words, which world is the *actual* actual world for that subject. A subject living on Twin Earth, for example, is not in a position to know any propositions involving the concept *water*; *a fortiori*, she is not in a position to know its application-conditionals. The crucial point, however, is that the capacity a subject on Earth would manifest in endorsing — and thereby coming to know — the application-conditionals of *water*, is just the same capacity that a subject on Twin Earth would manifest in endorsing the application-conditionals of *her* concept. That capacity is the capacity to pick out, for each possible world-description, the value of the epistemic intension of one's concept on the hypothesis that it is true of the actual world. Since the epistemic intension of *water* is just the same as the epistemic intension of the concept they express on Twin Earth by their word 'water', what it takes for a subject to possess the concept *water* is just the same as what it takes for a subject to possess the Twin Earth concept.

The concession, therefore, does not change the fact that possessing a concept is, according to two-dimensionalism, a matter of dispositions or capacities constitutively independent of one's environment. This is the aspect of the view I will be criticizing here.

Now, one may wonder whether *any* fixed set of non-trivially specifiable cognitive capacities — over and above, that is, the capacity to think thoughts involving the concept in question, and perhaps to know its homophonic satisfaction-conditions — are entailed by possession of any particular concept. This is the objection to the analyticity strategy I raised in Chapter 2, and it obviously appears damaging to the two-dimensionalist view. I do not, however, intend to repeat this particular objection in the present chapter. On the contrary, let us concede for the sake of the argument that there is a significant class of concepts possession of which is associated with some distinctive, non-trivially specifiable cognitive capacities — even if the association is looser than what proponents of the analyticity strategy typically suppose.

It is surely the point of many (though by no means of all) of our concepts that they pick out, and thus enable us to classify, things we encounter in our everyday experience. Concepts otherwise very different from each other — such as *yellow*, *round*, *mountain*, *cat* or *chair* — all seem to serve such a function.⁸ It is natural to think that, for a certain notion of normality, possessing these concepts is normally — even if not always or necessarily — associated with a capacity to correctly apply them in the course of one's experience. We should expect — and, intuitively, we *do* expect — that a subject who possesses the concept *yellow* will, under favorable conditions, be able to classify yellow things as yellow by sight, and that a subject who possesses the concept *mountain* will be able to classify mountains as mountains, given appropriate information. Finding a place for this natural thought in a plausible account of conceptual thought is, as we have seen, much trickier than proponents of the analyticity strategy suppose. Nevertheless, for the purposes of this chapter, I am going to make the assumption that there are some concepts possession of which normally entails a capacity to pick out their extension. My argument will be that two-dimensionalism *still* needs a very implausible conception of such capacities.

3. RECOGNITION AND INFERENCE

1. Consider a concept of this recognitional sort — for example, the concept *water*. Combining the points of the last section with the two-dimensionalist picture, we get the claim that possession of the concept *water* normally entails having all that one would need to be in a position to know its application-conditionals through reflection alone (contingent limitations aside — from now on I will avoid explicitly mentioning this qualification).

⁸ The class of concepts that do not — or do not primarily — serve this function is very diverse, of course. It includes, for example, normative concepts and logical concepts.

Now, in order for this claim to be even remotely plausible, we would need first to exclude cases where possession of the concept intuitively does not involve a reliable disposition to correctly apply the concept. We must, therefore, exclude subjects whose grasp of the concept is deferential, or who adhere to confused but consistent and not easily refutable theories about water, and the like. Assuming this to be possible, the most plausible candidate for a capacity normally entailed by possession of the concept is the capacity to recognize water by its manifest qualities in conditions ordinary for us — that is, the ability to recognize as water liquid water, snow, ice, steam and the like, at least in ordinary environments. Thus the two-dimensionalist claim must be that possessing this recognitional capacity entails being in a position to know the relevant application-conditionals through reflection alone.

How plausible is this claim? Clearly, the assumption that a subject possessing the concept *water* normally has the relevant recognitional capacity entails that, given a suitable form of access to a world or world-description, she will be able to pick out the water in it. This, however, is a conditional claim about what the subject will be in a position to do or know under certain specified conditions. By itself, it as yet tells us nothing about whether the subject is in a position to know the corresponding application-conditional at all, let alone know it by reflection alone. In order to assess this latter claim we need to be more precise about what we take the relevant recognitional capacity to consist in.

In the following couple of sections I will discuss an account of recognition which, if true, would support the two-dimensionalist claim (or at least a reasonable reformulation thereof). I will begin with a rather crude version of the account, which I will subsequently refine. Still, I believe that this whole approach to recognition is very implausible, and in fact evinces a commitment to a very unattractive picture of concept-application. After arguing against it in §3.3, I will sketch an

alternative account of recognition (in §3.4). That account has no tendency at all to support the two-dimensionalist claim.

2. Here is a crude picture of recognition, which nevertheless straightforwardly supports the two-dimensionalist claim. To recognize water in a given world-description is, on this picture, a matter of discursive reasoning.⁹ Given the relevant world-description *D*, our subject brings to mind a specification stored in memory of what water is like, couched in the same vocabulary as *D*. She then goes through *D* with this specification in mind, figuring out what, if anything, it applies to. Accordingly, to have a capacity to recognize water would consist precisely in having stored in memory a specification of what water is like, which can be recalled and compared with an appropriately described situation (in the case where the situation is directly experienced we can assume that recognition will work on an appropriately formulated description derived from the experience).¹⁰

It is clear that on this inferentialist picture the recognitional capacity is also a capacity to know the relevant application-conditionals. In fact, on this picture there is no real difference

⁹ I use the term “discursive reasoning” for reasoning that can be represented in words, regardless of whether the subject who engages in it consciously goes through some kind of internal monologue or not. I take it as obvious that not all reasoning is discursive in this sense. When considering how to get a large sofa down a narrow stairwell, for example, I am definitely engaged in a process of reasoning. But it is clear that I do not need to have any knowledge of the geometrical theory that would be required to formulate and solve this problem in a language. My reasoning, therefore, is not discursive in the present sense.

¹⁰ This account (as well as the refinement of it I will discuss shortly) suffers from an obvious difficulty, which for reasons of simplicity I will ignore. No specification involving only the qualitative properties of water can capture the required recognitional capacity, because no such specification discriminates between water and twin-water (a substance found only on Twin Earth, a distant but actual planet). But, intuitively, if a person from Earth calls twin-water ‘water’, she has *not* recognized twin-water as water: she has, rather, made a mistake. Her recognitional capacity has misfired in this case. The problem is that on the present account no mistake can be found in her performance, since she has correctly judged that her specification applies to twin-water.

What we need seems to be a requirement that the object picked out be the one the subject has actually interacted with in the past, and from which the stored specification derives. So the specification would end up looking like this (where *S* is supposed to be a specification involving only qualitative properties): “the *x* such that: (*x* is *S* & my past interactions with *x* are the source of *S*)”. These complications will be ignored in the text, since there are more fundamental problems facing the proposed account.

between recognizing water and reasoning by *modus ponens* from the appropriate application-conditional and world-description. This crude inferentialism about recognition, of course, is not often explicitly defended: it is extremely implausible — just on plain phenomenological grounds — that recognition must be a matter of discursive reasoning. We usually don't know accurate enough discursive specifications of the things we are able to recognize, and, even in the exceptional cases where we do, we certainly do not seem to use them in the way this account supposes. But even if the crude account fails, a more refined version of the same idea might still seem tempting.

Recognition is, obviously, driven by the way something looks (or, more generally, the way it appears to our senses). This might make it seem that recognition *must* be a matter of comparing a specification stored in memory with what is currently before one's senses. If the crude picture is misleading, therefore, one might suggest that the reason is that it assumes that the stored specification must be discursive. But this assumption is not compulsory: the stored specification might be something more like a mental image which the subject cannot fully put into words, for example.

How would this admission affect the two-dimensionalist claim that possessing the recognitional capacity puts a subject in a position to know the relevant application-conditionals through reflection alone? Unless the claim is reformulated in some way, it would seem that such an admission would undermine it. As Steven Yablo points out, a lot of our concepts — our “observational concepts”, as he collectively calls them — require for their application an exercise of our sensible capacities. Even in cases where the application is only hypothetical — cases in which we are trying to work out what the concept would apply to, on the hypothesis that such-and-such turns out to be the case — we need an “offline”, imaginative exercise of our sensibility.¹¹

¹¹ Yablo (2002: 441-92).

One of the cases Yablo discusses in detail is that of a certain family of oval figures, called “Cassinis”.¹² The shapes can be defined algebraically as the sets of points on the Cartesian plane satisfying equations of a certain form. But, as Yablo points out, knowing its algebraic equation is not enough to know that a Cassini is oval: the only way to know *that* is to actually plot the figure, either in one’s imagination or on a piece of paper. But neither of these methods, according to Yablo, qualifies as knowledge through reflection alone. Knowing that Cassinis are oval, as he puts it, requires “peeking”. Thus, even if it is true that a subject who possesses the concept *oval* (and the relevant mathematical sophistication) is in a position to know the conditional “if something is Cassini-shaped, then it is oval”, she is not in a position to know it *through reflection alone*.

This argument is convincing as far as it goes, but it doesn’t go far enough. There is no deep reason, I think, why an inferentialist about recognition cannot build “peeking” into her account.¹³ More generally, one could state the inferentialist view in a way that avoids commitment to any particular account of *how* the relevant specification is stored and compared to what is currently before one’s senses.

What the refined inferentialist account should require is just the following. Suppose that, in a specific experiential context, one knows a proposition by recognition (this proposition might be an identity proposition, such as “this is Fred”, or a proposition subsuming something under a kind-concept, such as “this is oval” — where “this” is a perceptual demonstrative). Then, one’s grounds for that knowledge must consist in the cogency of an argument, whose premises include: (i) a formulation in the appropriate phenomenal vocabulary of a specification stored in the subject’s memory, and (ii) a description of her current experience, formulated in the same phenomenal vocabulary.

¹² Yablo (2002: 467-9). As Yablo notes, not all Cassinis turn out to be oval; but, like Yablo, I ignore the point here.

¹³ This would seem to be Chalmers’s own response: “given the phenomenal information [...] about appearances (in a pure phenomenal vocabulary) ‘peeking’ comes for free”. See Chalmers (2002a: 185).

The crucial point here is that, in order for the grounds of one's recognitional knowledge to consist in the cogency of this argument, one need not actually arrive at the knowledge by thinking through the argument itself. Of course, in order for the inferentialist account to be true, recognition should involve on the part of the subject some form of ability to reason on the basis of her stored specification and her current experience to the recognitional proposition. However, the details of any particular process of recognition need not be relevant to the truth of the inferentialist account, as long as an argument of the above form captures its epistemic thrust. If this is correct, then two-dimensionalism should also be allowed to help itself to the same unspecified sort of process as the relevant way of coming to know the application-conditionals of one's concepts. As Chalmers puts it, peeking may come for free.

Nevertheless, as I will argue in the next section, we have reason to reject even this refined account: recognition is not to be construed as a matter of inference at all. To recognize an object is to non-inferentially identify an object presented to one with an object one has encountered in the past. To recognize an instance of a kind — such as oval or water — is, analogously, to non-inferentially identify something in one's environment as an instance of that kind. If this is right, then recognitional capacities offer no support to the two-dimensionalist claim, even in its more relaxed version.¹⁴

3. Consider again the case of the concept *oval*. I take it that *oval* is what we may call a “manifest shape” concept. It is a concept that picks out certain geometrical shapes, but it does so in virtue of a correlation with the way they appear to us. Thus *oval* is not to be identified with any concept formulated in exclusively mathematical terms (and not just because of vagueness). Suppose, now,

¹⁴ Even though Yablo does not distinguish between inferentialist and non-inferentialist accounts of recognition, some of the things he says suggest he has a non-inferentialist account of recognition in mind. See, primarily, the argument in Yablo (2002: 473-4; also 464-5). If his intention is indeed to argue against the inferentialist conception of recognition or concept-application in general, then his argument about peeking is not sufficient.

that possessing the concept *oval* consists in having the capacity to recognize oval-shaped things as such, and that recognition is conceived inferentially — recognizing something, that is, is a matter of matching its current appearance with a stored specification. Now, we can certainly imagine world-descriptions in which, for whatever reason, some things whose contours approximately satisfy circle equations systematically appear oval to us. We may suppose, along with Yablo, that a world-description in which dimes appear oval to us is true of the actual world. Would that be a case in which dimes turn out to be oval, instead of round?

It seems clear that it would not: the existence of such a glitch in our visual system should, intuitively, have no effect on the manifest shape of dimes. Whether dimes are oval or round depends only on their geometrical properties. If we accept the inferentialist account of recognition, however, it seems that we have to deny that. Since recognition is a matter of the fit between a stored specification and an appearance, on the hypothesis that the way dimes actually appear matches the specification, they are *genuinely recognized* as oval. Given the factiveness of recognition, dimes turn out to be oval.

Conversely, if we make the supposition that some oval-shaped things — for example, the dining-room table — do not look oval in the actual world, should we say that on that supposition they turn out not to be oval? Again, that seems false: in order for the table not to be oval it has to have different geometrical properties; it's not enough that the way it looks to us be different. Just as before, however, if recognizing something as *oval* consists in matching a specification with an appearance, we have to wonder whether we can avoid the conclusion that the table turns out not to be oval.¹⁵

¹⁵ On this point, I have encountered the objection that, whatever capacity a subject has to recognize manifest shapes (and oval among them), it is retained when she considers other hypothetical scenarios as actual. Thus, it is objected, when the subject considers the specified hypothesis as actual, she can tell that dimes are round and not oval, and accordingly she can discount her appearances on that supposition as misleading. As a point about the evaluation of conditional claims of the relevant form this appears to be all right: after all, we do

This might not be a problem if the subject had other resources, additional to her perceptual appearances and her stored specification, sufficient to rule out the false positives or negatives on the relevant hypotheses. I do not think the resources two-dimensionalism allows are in general sufficient for that, however. Consider how this would be supposed to work in the first of the above examples. In this case, the issue is to rule out the result that our subject recognizes something as oval when it is not, in fact, oval. Our subject has in her possession, we are supposing, a complete — matters of manifest shape aside — description of the world, including both the geometrical properties of things, and how shapes appear to her and to others. Could some of this collateral information tip her off that, despite the match between her appearance and the stored specification, she should not apply the concept *oval* to dimes?

If the distortion affects only a small subset of our subject's appearances (perhaps only approximately dime-shaped round things tend to appear as oval to her) then that very fact might alert her that something is going wrong. The subject, we may assume, knows that whether *oval* applies to an object or not depends only on its geometrical properties; since she knows that the geometry of dimes is similar to that of other things that appear to her round rather than oval, she might reason that her dime-appearances are distorted. Thus she would withhold her application of *oval* to dimes. Even if *all* of our subject's shape-appearances are systematically distorted, she might

intuitively judge that dimes are not oval on the relevant hypothesis. Nevertheless, the question here is to *account* for our ability to apply the concept oval, whether hypothetically or not (and thus also for our ability to evaluate the relevant conditionals). This is precisely the feature of two-dimensionalism we are in the process of evaluating. We cannot, therefore, simply *assume* that the ability to correctly apply the concept oval carries over to the hypothetical scenario. Instead, we must examine whether the resources available to two-dimensionalism really do amount to a capacity to correctly apply the concept hypothetically.

The point may be illustrated by considering a simple attempt to capture the epistemic intension of *oval* by a description. Consider the following biconditional: 'x is oval if and only if x looks oval in the actual world (to normal subjects and in normal conditions)'. Such a biconditional seems to be in the spirit of two-dimensionalism (general doubts about the coherence of introducing concepts by means of such biconditionals are not relevant here). This biconditional makes it clear that, on the hypothesis that dimes normally look oval in the actual world, dimes *are* oval. The analogous result might be right for so-called "response dependent" concepts, such as the concept *cool* (as in "cool haircut", not "cool weather"). But, intuitively, such concepts lack a kind of objectivity that concepts of manifest shape possess.

still be able to reason her way out of the problem as long as we assume that the distortion is statistically abnormal among her peers.

Consider, however, how things would be if we assumed instead that the glitch has a systematic effect on all of a statistically normal subject's visual perceptions. The way dimes appear to our subject in this case is not statistically abnormal at all: this is just how one would expect dimes to look to our subject, given how everything else looks both to her and to her peers. Now, what could tip our subject off that her dime-experience is not veridical? In order for our subject to figure this out, she must be capable of figuring out how dimes *ought* to appear. But how could she do such a thing? By hypothesis, actual appearances are not going to be of any help: as we have assumed, the way things appear to her is systematically distorted, so that round things appear to her oval instead of round. But neither would her otherwise complete description of the world suffice. For one thing, the description of the world available to her must be free of manifest shape concepts (otherwise the two-dimensionalist claim we are evaluating becomes trivial). But neither can she rely on the appearances of her peers, since they are — as we are assuming — systematically distorted in just the same way as her own. Since neither her appearances nor the description of the world she is considering can be of any help, our subject has no resources that would suffice for her to figure out that her dime-experiences are not veridical.

4. The kind of difficulty discussed in the last section does not depend on the specific features of the examples employed. It rather results from some central features of the two-dimensionalist position. In the next paragraph I will describe where I think the problems stem from. I will then suggest a different view of recognition, which avoids such difficulties.

The concepts I have been concerned with are associated with recognitional capacities. Given such a concept, it is always an a posteriori matter what (if anything) is in its extension — what, if anything, it applies to in the *actual* actual world. That is not something that can be settled by

reflection alone. This, I take it, is uncontroversial. According to two-dimensionalism, however, this further implies that the connection between the recognitional capacity and the concept's extension is not constitutive, but only contingent. This follows from the fact that, when determining the epistemic intension of a concept, the only things we are supposed to hold fixed across every possible world-description are those that can be settled by reflection alone. If, however, the recognitional capacity were constitutively linked to an extension, then that link, even though not knowable by reflection alone, would settle the value of the concept's epistemic intension for every world-description. This, I suggest, is why two-dimensionalism requires an inferentialist account of recognition. (In §3.5 I will provide an additional argument, to the effect that two-dimensionalism cannot shed this commitment.) On this account, recognition consists in matching a specification stored in one's memory with one's current appearances. Since it is surely contingent which things cause appearances that match the specification, the contingency two-dimensionalism requires is thereby secured. But this is also just what leads to the unacceptable results of the last section.

Accordingly, the problems are avoided if recognition is not a matter of matching a stored specification with an appearance.¹⁶ Denying the inferentialist model of recognition does not amount to denying the obvious fact that recognition is driven by the way an object appears to us, and therefore that it requires information about the object's appearance to be stored in us in some way. The point is simply that the process by means of which the stored specification is compared to the present experience is not to be construed as having the epistemic structure of an inference. This process may be, for all that the phenomenology of recognition requires, just a sub-personal, neural process, not to be construed as a cognitive act of the subject herself at all.

¹⁶ Strictly speaking, denying inferentialism is not the only way to avoid the problems of §3.3. There seems to be room for an externally constrained inferentialism, according to which what *oval* applies to is held constant, even though recognition still operates inferentially (the possibility of this hybrid view was suggested to me by John McDowell). This view is still not acceptable to two-dimensionalists for just the reasons explained in the text: it requires us to hold fixed something that is not knowable by reflection alone.

So what would the non-inferentialist account of recognition look like? Consider first the case of recognition of a particular object o . According to the present account, recognizing o in one's perceptual environment is to non-inferentially know a proposition of the form “this is o ” (where ‘this’ is an appropriate perceptual demonstrative mode of presentation of o).¹⁷ To have a capacity to recognize o will, accordingly, be to be able to non-inferentially know that o is identical to an object in one's perceptual environment, if it is presented to one under an appropriate appearance.

One thing to immediately note is that, on this view, recognitional knowledge does not depend on attributing any particular appearance to o . For all that the present account requires, the subject need not be able to think of the way o appears at all, except perhaps by using the very same perceptual demonstrative — as in “ o looks like *this*”. One derives propositions about how o appears from one's recognitional knowledge, rather than the other way round. (On the other hand, the specification of a recognitional capacity must make reference to a specific appearance. This reflects the intuition that, for example, two people who have known Fred in two completely different periods of his life — perhaps one knew him as a young child and the other as an old man — will have distinct recognitional capacities with respect to the same person.)

The case of recognition of kinds can be treated in an exactly parallel fashion. Suppose that ‘this’ expresses a demonstrative mode of presentation of an item in one's present perceptual context. To recognize that item as belonging to a kind k will be to non-inferentially know the proposition “this is a k ”. To have a capacity to recognize k will be to have the capacity to non-inferentially know of instances of k , presented to the subject under an appropriate appearance, that they are instances of k . Again, the way the kind looks to the subject is relevant to the individuation of the recognitional capacity, but the knowledge the subject acquires by recognizing instances of the kind is

¹⁷ I am assuming here that states of knowing are individuated more finely than reference is, so that the knowledge described in the text does not collapse to knowledge of the proposition that o is o .

not derived from prior knowledge that something with that appearance is present in her environment.

If this non-inferentialist account of recognition is correct, then the two-dimensionalist claim must be false. I will argue for this explicitly in the next section, but the reason why should be intuitively clear: on the present account, recognitional capacities are individuated in part in terms of what they pick out in the subject's environment — in the *actual* actual world. If, on the hypothesis that the world turns out to be otherwise, the subject happens to pick out something different, that is just a mistake: on that hypothesis, the subject's recognitional capacity simply misfires. According to two-dimensionalism, on the other hand, the extension of a concept cannot play a constitutive role in individuating the capacity that possessing that concept consists in. This, as we saw, is why two-dimensionalism leads to the unacceptable results of §3.3.

5. We have seen an argument against the inferentialist conception of recognition, which I suggested two-dimensionalism is committed to. I will now argue in a more detailed way that this commitment is not optional. I will do this by arguing that a hybrid view combining the main two-dimensionalist thesis with a non-inferentialist account of recognition does not result in a coherent account of concept-possession.

It might be suggested (although I do not think it actually *has* been suggested) that, even though there is no specification the subject employs in order to pick out what her concept applies to, there still is a specification which describes her non-inferential recognitional capacity. After all, assuming that the subject has such a capacity, one could define a function from ways the world may turn out to be to items of the appropriate category for the concept to apply to, such that for each way the world may turn out to be such an item is the value of the function if and only if it is what the subject would pick out if the world turned out to be that way. (I will call this function the 'recognition-function'.)

It is very questionable whether the disposition captured by the recognition-function can correctly be described as a disposition to *know* the application-conditionals, even assuming that it may capture a disposition to merely *conform* to them. The two-dimensionalist claim is that possessing a concept puts one in a position to know its application-conditionals through reflection alone. But if the capacity to apply the concept through recognition of its instances is non-inferential, then that capacity does not entail that the subject herself has any direct cognitive access to the conditions under which the concept applies. On the non-inferentialist picture, therefore, it is more plausible that a subject is only able to know the application-conditionals of her concepts a posteriori, through a course of experience in which those concepts are employed.

But this is not the objection I want to push here. An even more fundamental problem with the proposal to combine two-dimensionalism with a non-inferentialist account of recognition is that the proposed recognition-function cannot even capture a disposition to *conform* to the application-conditionals, on pain of rendering the two-dimensionalist account of concept possession incoherent. This argument gives us even more reason to think that two-dimensionalism must be committed to inferentialism about recognition in concept-application.

Now, if there is any plausibility in the idea that the recognition-function captures a disposition to conform to the application-conditionals for a given concept *C*, it must be that for every world-description, the value of the recognition-function for a subject who possesses *C* is the same as the value of *C*'s epistemic or primary intension for that world-description. But the epistemic intension of a concept fixes what it applies to for every world-description, on the supposition that the description is true of the actual world. Thus if the subject's recognition-function for a concept *C* is guaranteed to match its epistemic intension, then the subject is guaranteed to never be mistaken in her application of the concept.

This, however, is not a coherent account of concept possession. To possess a concept is to be subject to a normative standard for one's performances — this seems to be one of the few uncontroversial lessons of Wittgenstein's *Philosophical Investigations*.¹⁸ But we cannot make sense of a state's placing its subject under a normative standard if the state is simply *defined* so as to make every performance of the subject count as correct. This, however, is precisely what we have done by stipulating that the epistemic intension must match the subject's recognition-function.

To return to Yablo's example, there are possible world-descriptions in which a visual glitch makes our subject see dimes as oval. Supposing one of these to be true of the actual world, she will pick dimes out as items to which the concept *oval* applies. The intuitive response is that this will be a mistake on her part: dimes do not stop being round simply because our subject takes them to be oval. But on the present account no mistake has been made: the concept our subject uses has an epistemic intension that is guaranteed to match her recognition-function; thus, if on the hypothesis that the above world-description is true of the actual world she is disposed to pick dimes as falling under her concept, then dimes *do* fall under her concept on that hypothesis.

Perhaps one might object here that, in the spirit of two-dimensionalism, the recognition-function should not have been defined in terms of the subject's mere disposition to pick out a certain item, but rather in terms of her disposition to pick out an item *only after ideal reflection*. If the definition of the recognition-function is thus qualified, the response would go, then the identification of the recognition-function with the epistemic intension would be rendered harmless: since no actual subject is capable of ideal reflection, mistakes will be made just to that extent.

This response, however, simply misses the point. The account we have been discussing is supposed to incorporate the non-inferential account of recognition. If this is so, then the distinction between ideally reflective and ordinary concept-application cannot do the work it is supposed to do.

¹⁸ Wittgenstein (1958). The idea is also at the heart of Kripke (1982).

If the way a concept is applied in experience is non-inferential, then — intuitively — we must make room for mistakes that are not errors in reasoning of any sort. I take something to be oval, and I turn out to be wrong. Since my application of the concept is not the product of a process of reasoning, my error is not an error in reasoning either. It may be simply a “brute” error (as opposed to a “procedural” one) in the sense that, no matter how finely we analyze the process leading up to it, the error may not be due to anything that *I* did wrong at any particular step of the way. Since brute error is not due to anything the subject does wrong in the first place, granting her ideal powers of reflection should do nothing to shield her from brute error. It follows that even ideally reflective concept-application cannot be infallible. But since, on the present proposal, two-dimensionalism identifies a concept’s epistemic intension with what the subject applies the concept to under ideal reflection, ideally reflective concept-application *has* to be infallible. The possibility of error, therefore, is still not properly accounted for.

As I already said, I do not think this type of account has been explicitly defended by any proponent of two-dimensionalism. Perhaps Jackson comes closest, in a discussion of Gareth Evans’s non-inferentialist account of recognition. Jackson’s claim in that context is that recognitional capacities can be assimilated to capacities for descriptive identification, even though recognition itself remains non-inferential. According to Jackson, this is because “being disposed to be recognized by a subject as a such-and-such is a descriptive property of something that can and does serve to pick it out from other things”.¹⁹ Of course this is a descriptive property that can serve to pick out certain things; but it cannot be the property, or the concept, of a “such-and-such”, for in that case the subject would be rendered infallible in her judgments about which things the concept applies to: if she feels disposed to call something a such-and-such, then that very fact *makes* the thing

¹⁹ Jackson (1998: 66). See Evans (1982: 287-289).

a such-and-such. An attempt to qualify this by appeal to ideal reflection will fail, in the way just discussed.

6. One might perhaps accept the argument so far, and the non-inferentialist account of recognition that is its upshot, but still wonder whether the non-inferential character of recognition might be just an accident, due to contingent limitations of our cognitive apparatus. Perhaps we humans need non-inferential, hard-wired recognitional skills to overcome limitations of computational power: recognition in us works much faster and perhaps more reliably in the non-inferential fashion than it would if it had to rely on reasoning. But we can certainly imagine beings equipped with much more computational power. Perhaps such beings either have no need for non-inferential recognitional capacities, doing instead all the recognitional work through reasoning, or, while still relying in practice on non-inferential recognitional capacities, they are also capable of having personal-level attitudes to the information employed by their sub-personal recognitional systems.²⁰ It is arguable that such beings would be able to know the application-conditionals of their concepts through reflection alone.

Perhaps this may be thought not to matter, since it may be claimed that the concepts these beings use are different from the ones *we* use, due precisely to the different cognitive capacities used in their application. Suppose, for example, that these beings have a concept — call it *schmoval* — that classifies shapes using the very same appearances our concept *oval* uses. If this concept is used inferentially, then it will differ from our concept *oval* for the reasons outlined in §3.3 above: on the hypothesis that our beings' visual system has a glitch that makes dimes appear oval to them, dimes

²⁰ If concept-application is inferential all the way down, then, plausibly, there must be some basic concepts that are applied on the basis of only trivial inferences. Such inferences will consist in detaching from a homophonic statement of the satisfaction-conditions of a concept, on the basis of an experience described using the concept in question itself. (See also n. 5 above.)

will be *schmoval*, but they will still not be oval. Thus *schmoval* and *oval* are not the same concept, since their epistemic intensions differ.

It is not, however, essential to the spirit of the current proposal that the supposed beings apply their concepts in this simple inferential way. Perhaps some externally constrained form of inferentialism is true of these beings: perhaps *schmoval* always tracks the manifest shape oval by means of appearance, even in worlds where the associated recognitional capacity fails to pick that shape out. From this it follows that two-dimensionalism is still false of *schmoval*, since its epistemic intension is a constant, with its value fixed by the property it picks out in our beings' *actual* actual world (see also n. 16 above). Nevertheless, it may be worth considering the proposal for its own sake, even though it cannot help save the letter of two-dimensionalism.

Now, let us imagine beings of whom this externally constrained inferentialism is true, and who are moreover able to infer both ways between the algebraic equation a figure satisfies and the way it appears to them. They are able, just by looking at (or visually imagining) a plane figure, and with the help of background knowledge concerning their visual capacities, to infer to the algebraic equation it satisfies. Conversely, given an appropriate algebraic equation, they have no trouble inferring from it and the relevant knowledge about their visual capacities to what the corresponding figure would look like to them. Such creatures would not need to rely on non-inferential responses in order to group together the things we call oval: they would be able to achieve that classification inferentially, by means of explicit reasoning.²¹ Still, it does not seem altogether out of place to say that their concept *schmoval* is just the same concept as the concept *oval* that we use. After all, these concepts group together the very same shapes, and they do so by means of the very same

²¹ As already noted, at the most basic level this reasoning will simply involve applying a concept on the basis of its homophonic satisfaction-conditions and an experience described using that very concept. I am assuming here (as I have been doing throughout) that manifest shape concepts are not basic in this sense — that is, I assume that on the two-dimensionalist view there *will* be non-trivial application-conditionals that one must be in a position to know in order to possess these concepts.

appearance. But now, the argument goes, if such beings are indeed possible, couldn't we idealize away from the non-inferential character of *our* recognitional capacities, regarding it as just one more of the contingent limitations we in any case have to idealize away from?²²

I think this argument should be resisted. The kind of idealization proposed here is significantly different from the one that was built into the two-dimensionalist account from the start. That idealization involved abstracting from contingent limitations on information storage and processing capacities; crucially, it did not involve a fundamentally *new* source of information. But this is precisely what the present proposal involves. According to the non-inferentialist account of recognition (which is accepted as true of us by the present proposal), the information driving our recognitional capacities is not available to us through reflection; it is available to us only through ordinary empirical investigation. The proposed idealization invites us to lift that limitation. In effect, it invites us to consider what we *would* be in a position to know through reflection, if we suppose that a lot of information that, for us, is available only through empirical investigation, were available to us otherwise. There is no reason to think that the answer to this question is relevant to what *we* can and cannot know a priori.

7. For these reasons it is, I think, clear that we should reject the two-dimensionalist picture of conceptual thought, and the corresponding account of a priori knowledge. As I have been suggesting, it seems reasonable to attempt to ground the epistemology of the a priori in an account of conceptual thought. According to the two-dimensionalist version of this idea, the link is especially direct: possessing a concept just becomes a matter of being in a position to know by reflection alone a large number of non-trivial propositions. This, of course, is one version of the analyticity strategy — namely, the attempt to explain cases of non-trivial a priori knowledge by appeal to dispositions to judge or infer entailed by concept-possession. The argument against two-

²² This kind of response was suggested to me by Cian Dorr.

dimensionalism therefore reinforces the arguments against the analyticity strategy presented in Chapter 2.

As I already suggested in Chapter 2, however, these arguments do not affect the prospects of the more general project of giving an account of a priori knowledge based on an account of conceptual thought. I ended Chapter 2 by suggesting a starting point for such an account. In the following chapter I will attempt to complete that account.

Chapter 4
BASIC A PRIORI KNOWLEDGE AND KNOWLEDGE OF ONE'S OWN MIND

1. INTRODUCTION

In this chapter I will complete the positive account of a priori knowledge that has been my goal throughout this dissertation. So far I have argued for the starting point of this account. I began with the idea that, on a plausible view about the individuation of contents (or Fregean “thoughts”), contents are individuated in terms of the norms that govern the thinking of individual subjects or entire communities. But since contents are what may be the case, knowledge about their nature is knowledge about the shapes reality can take. Thus knowledge of the norms of thinking can lead to substantive knowledge — knowledge about what is or may be the case, and not just knowledge about how we happen to represent the world in thought or language.

In an earlier chapter I began to develop my approach by contrasting it to some of its closest relatives — namely, the views that follow what I called the “analyticity strategy”. Such views attempt to explain the possibility of a priori knowledge on the basis of dispositions to judge or infer that are supposedly entailed by concept-possession. As I argued, however, it is very implausible that possessing any particular concept entails any fixed set of dispositions to judge or infer; thus the analyticity strategy cannot get off the ground. Nevertheless, as I also argued, an epistemology of the a priori can be based on a more relaxed, holistic view of concept-possession, which does not require such entailments.

On this view, concept-possession is a matter of being subject — *normatively* subject — to the norms that serve to individuate the corresponding concepts. Normally, one is subject to such norms in virtue of participating in a conceptual practice. But one would not count as participating in such a practice — and thus as capable of drawing on its shared stock of concepts — unless one were, *on the whole*, sufficiently reliable in conforming to enough of the norms constitutive of the practice. This constraint is holistic, in the sense that there need not be any fixed set of cases that the subject must be disposed to get right in order to count as capable of thinking any particular thought, or as possessing any particular concept. Nevertheless, in the cases where the subject *is* in fact reliable, we can take her reliable dispositions as a starting point for an account of a priori knowledge.

As I also suggested, however, this by itself is not sufficient as an account of a priori knowledge. Simply being reliable in inferring in accordance with the law of double negation does not by itself guarantee *recognizing* its instances as good inferences (I am assuming, again, classical logic to be correct). One might be reliable in conforming to the law of double negation simply because of some random quirk in one's brain, while having no idea that such inferences are good ones. Intuitively, such “blind” dispositions cannot ground a priori knowledge. Thus, until we have a less schematic account of what it is for dispositions to judge or infer not to be blind in this way, we are still short of an epistemology of the a priori.

In §2 of this chapter I will try to make explicit the epistemological outlook that will inform my account. This will provide the framework for the substantive account, which I develop in the rest of the chapter. That account will be based on the idea (which I have already floated at various points) that, in order to fully realize the present approach to a priori knowledge, we need to connect it with a proper account of the self-awareness characteristic of conceptual thought.

2. THE QUESTION OF JUSTIFICATION

1. Once again, it will be helpful to begin by contrasting my approach to that of proponents of the analyticity strategy. Consider the question of how we can come to know a priori that the law of double negation is a valid rule of inference (assuming for the sake of the argument that classical logic is correct). As I suggested already in Chapter 2, it seems intuitive that coming to know this fact is a two-stage process. In the first stage one is able to conform one's thinking to particular instances of the law: if one is committed to $\sim\sim p$, in other words, one is thereby disposed to either commit to p or revise one's commitment to $\sim\sim p$. In a second stage, and given the required competence in meta-logical concepts, one recognizes that these patterns depend only on the logical form of the premise and the conclusion, and thus one can come to know that the law of double negation is a valid (truth-preserving) rule of inference. This two-stage picture — perhaps give or take a few details — seems to be generally accepted by proponents of the analyticity strategy as well.¹

Given this picture, the important question is how to understand the ability manifested in the first stage of our example. How are we to explain a subject's ability to conform her thinking to the law of double negation, independently of her already knowing that the law of double negation is a valid rule of inference? It is clear that not just any disposition to make transitions that match the law of double negation is going to be relevant here: a random quirk in our subject's brain might lead to such a disposition, but that would not count as our subject's being able to conform her thinking to the law of double negation. Being able to conform to the law of double negation is a *rational* ability, subject to questions of justification. The question, for the analyticity strategy as well as for the account I have been suggesting, is to give an account of the relevant kind of justification.

¹ Boghossian (2000); Peacocke (2003).

One typical move proponents of the analyticity strategy make here consists in appealing to the thesis that the relevant transitions are justified because they are “meaning-constituting” — because it is *in virtue of* being disposed to make the relevant transition that the subject possesses the concepts in question, or is capable of thinking the thoughts in question. Here is how Paul Boghossian puts the point:

If it is true that certain of our inferential dispositions fix what we mean by our logical words (in the language of thought), then it is very plausible that we should be entitled to act on those inferential dispositions prior to, and independently of, having supplied an explicit justification for the general claim that they are truth-preserving.²

Since the antecedent of this conditional is simply a statement of the central thesis of the analyticity strategy, Boghossian is happy to assert its consequent. A similar view is endorsed by Christopher Peacocke:

A thinker who meets the possession-condition for conjunction [i.e., a thinker who is disposed to find inferences of certain forms naturally compelling] does not need examples of an impossible kind that somehow determines [*sic*] a unique correct application in a new case. Nor does he need to have surveyed in advance all the correct applications. He just has to find certain transitions compelling in a certain way, and his finding them so has to have a particular explanation.³

Presumably, when Peacocke talks about what the thinker “needs” or “does not need” to do here he is talking about conditions of justification: if the subject is disposed to make the transitions mentioned in the possession-condition for conjunction, then nothing else is required in order for her to be justified in making those transitions.

² Boghossian (2000: 250). Boghossian’s use of the term “inferential disposition” is not without problems. Boghossian here is speaking of dispositions that are prior to — and indeed constitutive of — a subject’s understanding of the propositions involved. But it is questionable whether transitions that are not *the result of* understanding deserve to be called *inferences*. In what follows I will avoid Boghossian’s usage. In fact, criticizing Boghossian’s view on this point would be another way to press the same sort of concern that I will be pressing later on. (I thank John McDowell for pointing this out to me.)

³ Peacocke (1992: 13) Peacocke’s later view of implicit conceptions (2003) introduces further complexity into the picture, but it does not protect his view from the argument in the text, since he seems to positively envisage the possibility that the relevant states might be sub-personal, and thus in principle not accessible to reflection. On the other hand, since on the implicit conception view these dispositions are the *result of* understanding, this view is not — at least not directly — subject to the objection to Boghossian’s formulation mentioned in n. 2.

Now, even bracketing the arguments I gave in Chapter 2 against the idea that concept-possession entails a fixed set of dispositions to judge or infer, we should wonder whether this really is a good answer to the question of justification.⁴ Just why is it supposed to be the case that the fact that transitions that match certain patterns are meaning-constituting makes the *subject* justified in conforming her thinking to these patterns?

On the face of it, the possession of meaning-constituting dispositions seems at most to guarantee that the subject will not go wrong in her transitions; but it is not evident that this is sufficient for her to be justified in making those transitions. In general, it is doubtful whether the fact that a judgment or transition results from a reliable disposition is, by itself, enough to justify that judgment or transition, independently of the subject's own beliefs or other propositional attitudes. More specifically, in the case of a priori knowledge, I suggest that the basic capacity on which a priori knowledge depends is a capacity to conform one's thinking to a certain rule because one *recognizes* the correctness of its instances, rather than just a reliable disposition to make correct transitions. But the mere fact that certain dispositions are meaning-constituting for a subject — although sufficient for reliability — is not sufficient to guarantee that the subject recognizes the correctness of anything.

This line of criticism echoes Boghossian's own criticism of externalist views that would impose no condition on the relevant dispositions other than that they be reliably correct. As Boghossian observes, such views go wrong because they would count a subject justified in making a particular judgment or transition even though, intuitively, the subject's thinking is epistemically *irresponsible* on the corresponding occasion: there is nothing in the subject's thinking that justifies *her*

⁴ Of course there is a way of taking the idea of meaning-constituting disposition which would be all right, even in the context of the holistic view of concept possession suggested in Chapter 2. The argument against the view explicitly discussed in the text would apply to this view as well.

in trusting her dispositions — even though they are, in fact, reliable.⁵ In effect, what I am now suggesting is that Boghossian’s own positive view *still* fails to satisfy this requirement: if the subject does not know that her judgments or transitions result from meaning-constituting dispositions, how is the situation any different from simple reliabilism?

2. In the background of this criticism lies a thesis about epistemic justification which is, admittedly, quite controversial, and which Boghossian himself explicitly challenges. Although this is not the place to defend the thesis in its full generality, I will try to at least respond to Boghossian’s challenge.

The thesis is that, for a judgment or inference to be justified for a subject, it must be justified from the subject’s point of view: its justification must only depend on facts about it that are, at least normally, accessible to the subject upon reflection. There is an intuitive connection between rationality, responsibility and the subjective point of view that would seem to be violated if we allowed justification to depend on facts that are outside the subject’s ken.⁶

Now, there are ways of developing this thought that lead to overly strong, and — to my mind — implausible, views. But it is important to note that a rather weak form of this principle is both plausible in its own right, and sufficient to explain the inadequacy of the view that the mere fact that some dispositions are meaning-constituting justifies the judgments and transitions that issue from them. Suppose that a subject is justified in believing a proposition *p*. Suppose, further, that *q* deductively follows from *p*. The question is, then, under what conditions is the subject justified in

⁵ Boghossian (2003: 228).

⁶ This is almost equivalent to the thesis Boghossian calls “access internalism” (Boghossian (2003: 228)). The difference lies in that, while Boghossian’s access internalism requires facts about justification to be *always* available to the subject, I think it will do if we simply insist on such facts being *normally* available to the subject. Although this is not a topic I can discuss here at length, I doubt that the stronger kind of thesis is plausible for *any* kind of fact (see Williamson (2000: ch. 4)). But this should not discourage the idea that we have privileged access to *some* kinds of fact: it should, rather, make us more careful in spelling out what privileged access involves.

inferring q from p ?⁷ The thesis I want to defend is that, whatever those conditions turn out to be, they must be reflectively available to the subject herself. This is to be distinguished from a stronger — and I think unacceptable — thesis, namely the thesis that such justification must consist in facts which the subject can *use in a cogent argument* in support of her judgment or inference.

On the sort of view I want to defend here, a subject in a situation as described above is only justified in making the inference if she *recognizes* that p is conclusive reason for q (of course what this recognition consists in will have to be explained; this will be my main task in this chapter). But this recognition is not to be construed as a premise in an argument of the following form: “if x is conclusive reason for y , and I am justified in believing x , then I ought to believe y ; but p is conclusive reason for q , and I am justified in believing p ; therefore, I ought to believe q ”. Such an argument clearly appeals to the same ability to conform one’s thinking to the laws of deductively valid inference. Therefore, appealing to it at this point would be hopeless: as Boghossian notes, such an appeal faces a circularity problem akin to the one famously identified (in a different context) by Lewis Carroll.⁸ But there is no need to appeal to this — or any other — argument here, for this is

⁷ This question must be distinguished from the question of the justification the subject has for the *conclusion* — that is, for q itself. If all goes well, then the subject’s justification for q is just her justification for p : good deductive inference transmits justification from its premises to its conclusion (to use a useful term introduced in Wright (1985)). But in order for justification to be transmitted in this way, being justified in believing that p is not sufficient. If, for example, the subject believes that q on the basis of a completely mistaken view of the rational connection between it and p , we should not count her justification for believing p as successfully transmitting to her belief that q . The question in the text concerns precisely the conditions necessary for justification to transmit from the premise to the conclusion of a deductively valid inference.

⁸ The example Boghossian is working with is the following (2003: 225):

I’m in the mood for some music; what, I wonder, is on offer today in Carnegie Hall? A quick check of the schedule reveals that Martha Argerich is scheduled to play on the 20th. As a result, I come to believe that:

- (1) If today is the 20th, then Martha Argerich is playing today in Carnegie Hall.

A glance at the calendar reveals that in fact:

- (2) Today is the 20th.

With these two beliefs in place, I move immediately to the conclusion that:

- (3) Martha Argerich is playing today in Carnegie Hall.

not the role that recognizing the rational connection between p and q is supposed to play. The point is just the following: if I believe that q on the basis of an inference from p , then that entails that I believe that q , and I recognize that I do so for the reason that p . It is not at all clear why an argument concerning how I ought to match my first order beliefs to my beliefs about what is a reason for what might be thought to be relevant here.

Normally, one's recognition that a set of propositions R that one believes constitutes a reason for p makes a difference to what one believes *directly*: if I believe that R , and I recognize that R is a (strong enough) reason for p , I *thereby believe* that p . At least in non-pathological cases, no further process is relevant. For example, if I recognize that my firmly held beliefs that G. W. Bush is a Republican and that he is president of the United States together constitute good reason to believe that a Republican is president of the United States, I *thereby believe* that a Republican is president of the United States. No further process is required to put the latter belief in place, or in order for it to count as inferentially grounded in the former. Thus Boghossian's charge of Lewis Carroll-style circularity gets no traction.

Of course there may be pathological cases in which I have to do some work to convince myself of p , even though I both firmly believe every member of R and recognize that the members of R collectively constitute strong enough reason for p . If believing p would be very painful to me, for example, then I might need to do work in order to accept p , despite recognizing that I have strong reasons for it. Higher-level reasoning might be useful here — just as it might be if I were attempting to convince someone else of the correctness of an inference that she failed to accept. But such cases are clearly not the norm, and therefore they give us no reason to doubt that,

Boghossian — surely correctly — observes that it would be hopeless to attempt to give an epistemological account of this *modus ponens* inference by appeal to an argument of the following form: “(1) and (2) would offer conclusive reason for (3), and I am justified in believing (1) and (2); but if x is conclusive reason for y , and I am justified in believing x , then I ought to believe y ; therefore, I ought to believe (3)” (Boghossian (2003: 233-235)). Clearly, the sort of ability manifested in this latter argument is just the ability to conform one's thinking to the rule of *modus ponens* that we were seeking to explain in the first place.

normally, recognizing that R is strong enough reason for p is all that is required in order to believe that p for the reason that R .

Is there something inherently implausible about the claim that a subject can recognize the rational connection between — for example — $\sim\sim p$ and p ? Boghossian in a footnote suggests, without explaining why, that “rational insight” into facts about justification is “unimaginable”.⁹ If Boghossian has in mind a special-purpose faculty that is not part of the basic structure of conceptual thought, then perhaps he is right to simply dismiss the idea. But as I am going to argue, an ability to recognize facts about justification follows from some fundamental features of conceptual thought. And, in any case, in ordinary experience it surely seems that we can very often tell what is a reason for what, just on the basis of understanding the relevant propositions.

Of course doubts about this view cannot be dispelled until we have a positive view of what the ability to recognize such facts consists in. What matters for now is that, at least *prima facie*, it does not seem implausible, and it clearly satisfies the reflective availability constraint — at least if that constraint is not interpreted as requiring that justification must consist in facts the subject can use as premises in a cogent argument.

3. OVERVIEW

1. So how are we to give an account of our ability to conform our thinking to norms constitutive of our concepts that respects the constraint of reflective availability? The holistic constraint on concept-possession I defended in an earlier chapter explains why subjects conform to enough of the norms constitutive of their concepts, but — like the analogous stronger claim of the analyticity strategy — it does not yet show this conformity to be due to any sort of recognition on the subjects’

⁹ Boghossian (2003: 236, n. 16).

part. I suggest, however, that this is only because we have so far not taken into account another fundamental feature of conceptual thought — namely, its characteristic self-awareness. If we combine an account of the self-awareness of conceptual thought with the holistic constraint, we do get an account of our capacity to recognize instances of the norms constitutive of our concepts.

In a nutshell, the idea is as follows. Part of what our rationality consists in is that we normally are in a position to know both what we believe or judge, and the reasons why we do so. This is not to say that our first-order judgments always need to be accompanied by second-order *beliefs* about our reasons (although they often are). The point is, rather, that belief and judgment themselves are “reflexive”, in a sense that I will explain later on.

Of course, there is no guarantee in any specific case that our beliefs are based on good grounds: a subject’s reasons for any particular belief can be, objectively speaking, as bad as we like. Thus a subject may be aware (in the relevant, non-second-order sense) that she believes that p for the reason that R (a set of propositions), and thus (by the same token) take it that p follows from R , while failing to be in a position to *know* that p follows from R — because p does not in fact follow from R .¹⁰

However, the holistic constraint above entails that, on the whole, a subject’s pattern of belief must conform to enough of the norms constitutive of her concepts. Thus it must be the case that, in a sufficient number of cases, when our subject believes that p for the reason that R , that belief exemplifies a norm that is indeed constitutive of the concepts involved. Moreover, given the point about self-awareness, the subject is aware (in the relevant, non-second-order sense) that she believes that p for the reason that R , and thus she takes p to follow from R . But in these cases she is in fact reliably correct in taking p to follow from R . Thus, in these cases, her believing that p for the reason

¹⁰ Of course, it may happen that the subject’s awareness that her belief that p is based on R leads her to retract her belief rather than judge that p follows from R . But in this case the subject no longer believes that p for the reason that R . Thus such cases do not disturb the general point; if anything, they reinforce the connection between rationality and self-awareness.

that R is, intelligibly, describable as based on her *recognizing* that p follows from R . If she then goes on to *judge* that p follows from R , then that judgment can count as knowledge, based on her recognition of the normative relations among contentful states. Judgments of this sort constitute our basic a priori knowledge.

This, of course, is not to say that *all* knowledge of what is a reason for what is a priori. Consider the following example. I hear voices behind a closed door, and I conclude (truly, let's say) that there is a class going on behind the door. I know I believe that there is a class going on behind the door, and I take my perceptually grounded belief that there are people talking behind the door to be good reason for this belief. But this is only correct if I have further empirical knowledge — for example, that the room behind the door is a classroom, that today is not a school holiday, that the time is not past regular school hours, and so on. Outside this context, it does not follow from my belief that there are people talking behind the door that there is a class going on, and thus I cannot know that it does. The reason provided by the perceptual belief is, in terms borrowed from Crispin Wright, *information dependent*. In general, my belief in a set of propositions R gives me information dependent reason for a proposition p just in case: (i) R gives me a reason for p and (ii) there is a set of background propositions I that I believe and which is such that, were I not to believe some member of I , R would no longer give me a reason for p (at least not without other compensatory changes in my total cognitive state).¹¹ No doubt this formulation is vague at various points; but it surely captures a genuine phenomenon. The crucial point for present purposes is just the following: if, in a case of information dependence, I do know what is a reason for what, my knowledge may *still* not count as a priori, since it is partly grounded in background knowledge, which may (as in the case just described) be empirical.

¹¹ Wright (2003: 58-59).

Such cases contrast with cases where what is a reason for what is determined by norms constitutive of the contents involved. There is, of course, a sense in which even in the case where I infer p from $\sim\sim p$ I must rely on background knowledge: for, in order to infer p from $\sim\sim p$, my reasons for $\sim\sim p$ must be stronger than any independent reasons I might have to doubt p . But this case is still not at all similar to the one discussed above. Here the role of the background knowledge is not to secure that p follows from $\sim\sim p$, or that I know that it does. The nature of negation alone guarantees that if I believe $\sim\sim p$ then I had better believe p , or reconsider my belief in $\sim\sim p$.¹² If I am reliable in conforming to this rule, and if — on the basis of the self-awareness of belief — I come to judge that p follows from $\sim\sim p$, my judgment is not based on any potentially empirical background knowledge. Thus, it is basic a priori knowledge.

A lot needs to be said about every step of such an account, of course. Most of the rest of this chapter is devoted to this task. Before getting to the details, in the next section I will make a few comments about its general structure.

2. First, note that there is no ambition here to provide a list of informative necessary and sufficient conditions for a subject to be in a state of recognizing the normative relations among contentful states. The goal is to provide an account of some of the capacities involved in conceptual thought which will make it intelligible that, normally, subjects are in a position to recognize such things.

The situation is not unlike what we find in other cases of non-inferential knowledge, such as perception. An account of perception, I think, can only hope to explain why it is the case that, normally, if it perceptually appears to me that p , then I *perceive* — and thus am in a position to know — that p . It does not look plausible, however, that we can informatively spell out a full set of

¹² If one were inclined, in a Quinean spirit, to expect a reduction of rational relations to conditional dispositions to assent or dissent, then one would naturally be very suspicious of the distinction between rational connections secured by the nature of concepts and rational connections secured *via* background knowledge. But the account of content suggested earlier in this dissertation is not congenial to such a project; nor is there independent reason to commit to it.

necessary and sufficient conditions for a case of having it appear to one as if *p* to be a case of *perceiving* that *p*. At this point — in the epistemology of perception just as much as in the epistemology of the a priori — we need to appeal to intuitive judgments about interpretation. Our attribution to a subject of states of knowledge has to be part of a total interpretation of her cognitive state that makes sense of her interactions with her environment; such attributions, therefore, will be sensitive to a huge number of factors, which we cannot hope to spell out in a set of necessary and sufficient conditions.

This might prompt the worry that the present approach is just a cosmetic improvement on the views criticized earlier: the present view, one might object, does not respect the reflective availability constraint — it only seems to do so. We can begin by considering the following rough attempt to state the worry. If it is admitted that no necessary and sufficient conditions for a priori recognition are forthcoming, it follows that no necessary and sufficient conditions *the subject herself* can apply are forthcoming. But if the subject has no criterion that settles it whether she is in a genuine recognitional state, rather than in a state of mere *seeming* recognition, isn't the reflective availability constraint violated?

The worry as stated is confused. On the present view, it is the recognitional state itself, rather than *knowing that* one is in that state, that justifies one's basic a priori knowledge. If the relevant sort of justification were inferential, then the lack of a criterion (a set of necessary and sufficient conditions) that a subject could apply in order to distinguish genuine recognitional states from states of merely seeming to recognize *would* damage the prospects of recognitional states playing a role in justification. But since the relevant sort of justification is not inferential, it is not clear that the lack of such a criterion is a problem. The situation is again instructively analogous to the perceptual case. If perceptual justification were inferential, then my seeing that the sun is shining could not justify my perceptual knowledge that the sun is shining, since I normally have no

independent criterion — i.e., no criterion that I can apply without already knowing that the sun is shining — to guarantee the premise that I can see that the sun is shining. Any argument that relied on this premise would beg the question. But perceptual knowledge is not inferential, and therefore this problem does not arise.¹³

In order to respond to the worry in its more general form, it may be useful to compare the present view to “intuitional” views of basic a priori knowledge — for example, the view developed by George Bealer.¹⁴ Such views might, I think, be open to the charge that they merely seem to conform to the reflective availability constraint, while in reality violating it. According to such views, our basic a priori justification derives from “intuitions”, or intellectual — as opposed to sensory or, more broadly, imagistic — seemings: most people, for example, would be said to have the intuition that $1+1=2$, although of course they do not picture that $1+1=2$ (whatever that might mean). Crucially, on this view, intuitions provide *prima facie*, defeasible support for the propositions intuited. Assuming, for example, that one has an intuition that it cannot be the case that $\sim\sim p$ while it is not the case that p , one is thereby *prima facie* justified in believing that p follows from $\sim\sim p$.¹⁵

Now, on the intuitional view, it is clear that one may have an intuition that p even if p is not the case: there is no intrinsic connection between it seeming to one in intuition that things are a certain way and things being that way. But then, it seems reasonable to think that the *mere fact* that one has an intuition that p cannot be what really justifies one in conforming one’s thinking to p , independently of the specific credentials of that intuition. Bealer, of course, does provide an

¹³ The view of perceptual knowledge sketched here follows McDowell (1998a). This is not to deny that normally we *do* know that we are in a perceptual or recognitional state; it is only to deny that this knowledge needs to be independent of the corresponding perceptual or a priori knowledge. I will make some remarks about this kind of self-knowledge near the end of this section.

¹⁴ Bealer (1999; 2002).

¹⁵ Putting the point in this way would seem to leave Bealer’s view open to Boghossian’s charge of Lewis Carroll-style circularity: for it is not clear how one would apply one’s intuition that p follows from $\sim\sim p$ to any particular case, unless one also intuited that the particular case falls under the general rule. Bealer does not (to my knowledge) discuss the point, but I think an answer along the lines suggested earlier would be available to him.

account of what the credentials of particular intuitions consist in. In a nutshell, the view is that intuitions are veridical insofar as they result from “determinate understanding” — a mode of concept-possession that goes beyond ordinary competence.¹⁶ But then, it is hard to avoid the impression that what *really* does the justificatory work on Bealer’s view are not the intuitions, but rather facts about their aetiology: it is because intuitions tend to be produced by states of determinate understanding that they give one *prima facie* justification to conform one’s thinking to them. Intuitions are mere symptoms of the facts that do the real justificatory work. But, of course, these latter facts — facts about the aetiology of one’s intuitions — are not normally reflectively available to one. Thus the intuitional view would seem to only superficially respect the reflective availability constraint, while in reality violating it.

The view I propose here, on the contrary, is not open to this objection. The relevant conscious justificatory states on this view are supposed to be states of recognition. But states of recognition are *factive*, and thus we have no reason to doubt that they can play the justificatory role assigned to them.¹⁷

Admittedly, it remains fair to request of a proponent of an account of a priori knowledge along the present lines to offer an account of *how* we know that we are in a state of genuine recognition. The analogous request can fairly be made of proponents of accounts of perceptual knowledge along the parallel lines. The point of the above considerations was simply to make it clear that this request — fair though it may be — is a request for something *extra*, over and above the immediate concerns of an account of basic a priori or perceptual justification. There is,

¹⁶ Bealer (1999: 45; 2002: 101-107). I criticized Bealer’s introduction of the notion of determinate understanding in Chapter 2. For present purposes, I will not challenge the notion itself, but only question whether it can help with the particular epistemological problems at issue here.

¹⁷ For the role of factive states in knowledge in general see McDowell (1998a; 1998d) and Williamson (2000: ch. 1-3).

moreover, no reason to think that this request cannot be met. I cannot here offer a full discussion of the many delicate details of this issue, but I will offer a suggestion.

What seems to make the situation difficult is that, as we have agreed, the subject lacks a criterion that would allow her to distinguish between a genuine recognitional or perceptual state, and a state of merely *seeming* to recognize or perceive. But I think we have no reason to accept that such knowledge would *have* to be based on the application of such a criterion (either explicitly or implicitly). On the contrary, I suggest that one's knowledge that one *recognizes* a rational connection, or that one *perceives* that *p*, is simply derivative of one's having the corresponding ground-level perceptual or a priori knowledge. This is not to say that the self-knowledge here is based on the cogency of an argument from the perceptual or a priori knowledge. Just as there is no need to cash out in terms of inference the sense in which one's perceptual knowledge that *p* is based on one's perceiving that *p*, there is no need to cash out in terms of inference the sense in which one's knowledge that one perceives that *p* is derivative of one's perceptual knowledge that *p*. One's knowledge that one perceives that *p*, in a situation where one perceptually knows that *p*, is non-inferential, but it is attributed to the subject *on just the same basis* as the ground-level perceptual knowledge is. Similarly, in the case of a priori knowledge, one's knowledge that one is in a genuine recognitional state is non-inferential, but it is attributed to the subject on just the same basis as the corresponding basic a priori knowledge is.

A fuller treatment of this topic falls outside the scope of the current project. I think that a proper account of our capacity to acquire knowledge by perception shows it to involve not just the reliability of the subject's sensory faculties, but moreover enough capacity for reflection to make it intelligible that, normally, when we attribute to her perceptual knowledge in such a situation we

should also take her to be in a position to know that she has that perceptual knowledge.¹⁸ Similarly the account of basic a priori knowledge developed here should help make it intelligible that, normally, when we take a subject to genuinely recognize the normative relations between two contentful states, we should also be prepared to take her to be in a position to know that she does so (I make some further remarks about this in n. 29 below).

Let us now turn to the details of the connection between basic a priori knowledge and the self-awareness characteristic of conceptual thought.

4. BASIC A PRIORI KNOWLEDGE AND SELF-AWARENESS

1. We are normally in a position to know what we think. This knowledge is surely non-inferential. But it is even more direct than that: it is, to use a term loosely adapted from Anscombe, non-observational. For our purposes, we can define non-observational knowledge as a kind of non-inferential knowledge that is, in a sense, inseparable from the fact known.¹⁹ By this I do not mean that in all possible worlds where the fact obtains it is also available to be known; I mean, rather, that *if* it is available to be known, then such knowledge is (epistemically) grounded in nothing further than the fact itself.²⁰ This contrasts with observational non-inferential knowledge, which, for the

¹⁸ Williamson (2000: ch. 5) argues that there must be cases where one knows that p while failing to be in a position to know that p . His example concerns perceptual knowledge. Nevertheless, although Williamson's argument is persuasive against the thesis that, necessarily, if one knows that p one is in a position to know that one knows that p , it does not establish that cases where one lacks self-knowledge are not exceptions, rather than the norm. I think there are good intuitive and theoretical reasons to believe that this is the case (I discuss this point, and how it can be made compatible with Williamson's argument, in my "Apriority, Contingency, and Knowledge from Fallible Sources"). Given the holistic nature of attributions of mental states in general, and states of knowledge in particular, we should in any case not expect exceptionless generalizations in the area.

¹⁹ This definition would match most of the central uses of the term in Anscombe (2000). It would not, however, match them all. But Anscombe exegesis is not to the point here.

²⁰ Williamson (2000, ch. 4) argues that only trivial conditions are "luminous", in the sense that, necessarily, if they obtain we are in a position to know that they do. What I say in the text does not contradict Williamson's

purposes of the present distinction, can be understood as knowledge (partly) grounded in some sort of contingent epistemic relation between the subject and the fact known. In the paradigm case this relation is sensory perception (although memory or testimony may also be examples): if my knowledge that p is by visual perception, for example, then it is grounded in my *seeing* that p , not just in the fact that p . Our knowledge of our own minds is in this respect characteristically direct or unmediated, as it is sometimes put.²¹

If conceptual thinking is open to non-observational knowledge, then there must be something about its nature that explains why having a thought normally puts one in a position to know that one is having that thought. In what follows I will suggest an explanation based on the centrality to conceptual thinking of the act of judgment, or of making up one's mind what to believe.²² Judgment, as I will argue, requires that our beliefs be essentially *reflexive*, in the sense that believing that p partly consists in taking oneself to do so, and for certain reasons.

point: my claim is that there is a kind of knowledge such that, for any bit of knowledge of that kind, it is grounded in nothing more than that the fact known obtains. I do not claim that, necessarily, such facts can be known to obtain whenever they do. Take Williamson's own example, namely, the condition of feeling cold (2000: 96-8). Williamson argues that one is not in a position to know that one feels cold whenever one feels cold. This, however, does not contradict the claim that when one does know that one feels cold, one's knowledge is grounded in nothing other than just feeling cold. In fact, it is hard to see what *else* it could be grounded in.

²¹ This is not to deny that there are features of our mental states that we *do* need observation to know about. Ernest Sosa invites us to consider cases where we consciously represent complex images or entertain complex propositions. He then points out that some of that complexity may not be available to us, except through some kind of observation (BonJour (2003: 122-134)). One of his examples concerns representing arrays of dots. Suppose we have two arrays, of which the first contains, say, 11 dots, while the second contains 12. It is quite plausible that most subjects would fail to distinguish between their visual representations of these two arrays without counting the dots. But if one has to count, then one does not know in the direct, unmediated, non-observational way I have been describing. Such cases do not disturb my main point, since they have no tendency to show that non-observational self-knowledge does not exist. Where I disagree with Sosa is in his taking these cases as paradigms, around which an account of self-knowledge is to be constructed. What Sosa's account seems to me to miss is what I argue for in §7: self-knowledge should not be construed as a near-perfectly reliable add-on to our capacity to judge, but rather as an essential feature of that capacity itself.

²² Judgment in this sense does not have to be a *conscious* act of making up one's mind: a lot of our judging happens without our consciously paying attention to it. On the other hand, I also do not want to commit to taking *any* belief, no matter how acquired, to be the product of judgment. We can imagine scenarios in which one is completely passive in relation to some of one's beliefs — perhaps they are implanted by means of a surgical manipulation of one's brain, for instance. Beliefs acquired in this way are not the products of

What exactly this ‘taking’ consists in will have to be explained further as the account develops. For present purposes, I will just note two things. First, this ‘taking’ is not as such a second-order judgment: a subject may count as taking herself to believe that the sun is shining because she can see that it is, even though she does not *believe* that she believes that the sun is shining (even if there is something incoherent about believing that *p* without believing that one believes that *p*, that is not a consequence of my account of self-awareness). Nevertheless, this ‘taking’ is the backdrop against which full-blown cases of second-order belief and self-knowledge must be understood. The second thing to note is that by taking oneself to believe that *p* for the reason that *R*, one thereby also takes it that *R* objectively *is* a reason for *p*. This commitment is not a *belief* about what is a reason for what; it is, rather, one aspect of the self-awareness involved in the first-order belief itself. This aspect of self-awareness is essential to our capacity for a priori knowledge.

2. But why should we take the capacity to judge to require reflexivity in this sense? Intuitively, the capacity to judge is the capacity to make up one’s mind what to believe: it is the capacity to review, revise and add to one’s cognitive state.²³ Now, my claim is that one cannot possess a capacity to judge understood along these lines, unless one is — normally, at least — aware of one’s own beliefs and the way they hang together in an interlocking and mutually supporting system. Believing that *p*

judgment. Such beliefs, moreover, may fail to be available to non-observational knowledge; but such pathologies are not to the point here.

²³ This should not be taken to imply that judgment is a kind of intentional action, with knowledge (or perhaps just truth) as the subject’s intended aim: this would be hopeless for many reasons, not least because it would seem to require us to take intention to be intelligible independently of belief or judgment. There is, nevertheless, a sense in which we *can* say that knowledge is the “constitutive aim” of judgment; but what this should be taken to mean, I suggest, is that in attributing a capacity to judge to a being, we are constrained to interpret its total cognitive state in a way that, on the whole, maximizes knowledge.

A more traditional constraint on the attribution of judgment or belief to a being involves the maximization of truth, rather than knowledge (see, for example, Davidson (1974, 1983)). For a defense of knowledge, rather than truth, as the constitutive aim of judgment or belief see Williamson (2004). It is worth pointing out that even defenders of the former idea should be willing to grant the point I want to get out of this characterization of judgment — namely, that a capacity to judge requires a certain form of rational control over one’s total system of commitments.

involves *taking* oneself to believe that p , both in that one must recognize p to be part of the commitments one will rely upon in the future, and in that one must hold oneself responsible for reassessing one's commitment to p in the light of further judgments.²⁴

The starting point here is that judgment, or making up our mind, is our way of exercising rational control over our mental state. We cannot make sense of the idea that a being possesses a capacity to judge, unless that being is capable of adjusting some of its cognitive states in the light of other such — actual or possible — states. Spelling out what this requires, in turn, shows judgment to essentially depend on reflexivity.

Consider the following elementary piece of reasoning. Fred believes that Mary will be at a certain meeting tonight. In the meantime, however, he learns that Mary is sick with a fever. Accordingly, he revises his old belief that Mary will be at the meeting, and now believes that she will not be there. Even such elementary reasoning, I suggest, is unintelligible independently of the reflexivity of belief.

First, it is clear that Fred does not simply *add* to his mental state the belief that Mary will not be at the meeting (not, at least, in a non-pathological case). What happens is that Fred *revises* an older belief, in the light of new information that undermines its grounds. This shows, at the very least, that Fred must have some capacity to assess the grounds for his beliefs, and to replace beliefs that are no longer justified with others that are. This is not yet enough to show that Fred's first-order capacity for belief is *itself* reflexive, since his capacity to assess his beliefs might be a separate, second-order faculty, which periodically scans his first order beliefs for weak points and inconsistencies. If that were so, then neither his original belief that Mary will be at the meeting, nor

²⁴ As the formulation in the text makes clear, one should not conceive of the reflexivity of the capacity to judge atomistically: the reflexivity of the capacity to judge concerns the way in which judgments form an interlocking, mutually supporting system. Incidentally, this serves indirectly to respond to a challenge raised by Sosa (see BonJour (2003: 135-136)). The challenge is to explain our knowledge that we *lack* a certain belief. Since the reflexivity of the capacity to judge concerns the meshing of our judgments into a system, it can also serve as the basis of our knowledge that a certain belief is not part of that system.

the later judgment that she will not, need themselves involve self-awareness. Fred detects their presence, and their incompatibility, through the employment of his second-order faculty.

I do not wish to deny that we may indeed possess such a second-order faculty, and occasionally use it in just this way. But it is extremely implausible that this is what is involved in Fred's reasoning. For, on this picture, Fred *first* judges that Mary will not be at the meeting, and only *then* — in virtue of an exercise of his second-order faculty of assessing his first-order beliefs — does he give up the belief that Mary will be at the meeting. But this cannot be right: intuitively, Fred *never* believed both that Mary will be at the meeting and that Mary will not be at the meeting. But even if one insisted that the operation of the second-order faculty is instantaneous, so that there need not be any temporal interval during which both beliefs are present in Fred's mind, the point remains that, on this view, we explain Fred's reasoning by attributing to him belief in — or, if beliefs (as states) cannot be instantaneous, an instantaneous belief-like relation to — a contradictory pair of propositions, and the ability to remove the contradiction. But this would seem too complex a picture of what is an elementary piece of reasoning.

I think it is more plausible to hold that Fred's judgment that Mary will not be at the meeting *just is* his giving up of the belief that Mary will be at the meeting (excluding, again, pathological cases). To put it in a slightly metaphorical way, the first-order judgment itself involves awareness of the place the new belief must rationally occupy in Fred's web of belief, and makes room for it accordingly. When Fred originally believes that Mary will be at the meeting, his belief involves his being aware both that he believes that Mary will be at the meeting, and also that this opinion is grounded in a number of other considerations — including perhaps that she said she would be, that she seemed to be healthy enough yesterday, and so on.²⁵ When Fred finds out that Mary is sick with

²⁵ More minimally, Fred may just be aware that any one among a number of considerations will speak *against* Mary's being at the meeting. This point does not make a difference for anything in the text, as it still entails that Fred's beliefs involve his being aware of rational relations.

a fever, this directly undermines the belief that Mary will be at the meeting, since it directly undermines some of the considerations Fred acknowledges as its grounds. Thus his abandoning the belief that Mary will be at the meeting and his adopting the belief that Mary will not be at the meeting are just two sides of the same coin: they are not two distinct mental acts. Clearly, what makes this reconstruction of Fred's reasoning possible is taking his first order beliefs to be reflexive.

3. None of the above is meant to deny that there can be a system in some ways functionally similar to a full-blown cognitive system like ours, but which does not involve such reflexivity. We can imagine a system whose output is controlled by the contents of certain specially designated registers — its “belief-box” — and whose input is used in such a way as to simultaneously maximize certain desirable parameters characterizing the contents of the belief-box. If, for example, one of the contents in the system's belief-box is a representation of the proposition that the table is red, but subsequent input makes the proposition that the table is orange more likely to be true, then — other things being equal — the representation of the first proposition is replaced by a representation of the second, and the rest of the contents in the belief-box are modified accordingly. Such a system evidently displays some form of control, and even *rational* control, over the contents of its belief-box. But, I suggest, it still does not manifest a capacity to judge, because — having left reflexivity out of the picture — the control it displays is not of the right kind.

Of course something goes out of the belief-box and something else comes in, in accordance with certain rational rules; but that is not something our system *does*, in the sense of ‘doing’ involved in expressions such as “making up one's mind” that illustrate the relevant conception of judgment. Unless our system takes itself to be committed to the content that the table is red, and subsequently takes itself to be committed to the content that the table is orange instead (and thus to making the corresponding adjustments elsewhere), it is not properly reassessing and revising the contents of its

belief-box. To the extent that our system lacks a special perspective on its own contents, it lacks a certain kind of subjectivity required by a genuine capacity to judge.

We can make all this a bit more precise, in terms of an intuitive necessary condition of possessing a capacity to judge. In order for someone to possess a capacity to judge, some of her states that we are otherwise inclined to count as beliefs must be open to a special kind of explanation. Suppose that $F(S,X)$ is a schematic relational predicate, signifying that a subject S stands in a relation either to a proposition or a set of propositions that we are inclined to count as belief. Then, for some such relations, statements of the following form must be true:

$$F(S,p) \text{ because } F(S,R)^{26}$$

The ‘because’ in this schema clearly does not designate just any causal relation. If an instance of our schema is true of a subject, then it must specify our subject’s *reason for* believing that p . Consider, for example, the following intuitive instance of the schema: “Fred believes that Mary will not be at the meeting because he believes that she is sick with a fever”. The crucial point is that, when we take the latter case to be an instance of our schema, Fred’s attitude towards his own attitude and its explanation plays a central role. If it is true that Fred believes that Mary will not be at the meeting because he believes she is sick with a fever, then — pathological cases aside — Fred himself must be disposed to accept both that he believes that Mary will not be at the meeting, and that he does so because he believes that she is sick with a fever. If he is not disposed to accept the explanation, then this is serious grounds to suspect that the explanation is not true of Fred after all.

²⁶ This way of making the point follows Rödl (2007: ch. 3). More generally, the discussion in the text echoes discussions of self-awareness in action, inspired by Anscombe’s account.

One might worry that the schema in the text is unduly restrictive, since many of our beliefs seem to be justified otherwise than by other beliefs: my belief that the sun is shining, for example, is justified by my perceptual state, rather than by another belief. It is true that a creature which had *only* beliefs justified by contentful states other than beliefs would seem to be ruled out, without argument, by the formulation in the text. But, since I think a creature of this sort seems intuitively unlikely, I would rather not mess with the simplicity of the principle in the text just in order to make room for it. The point of the principle in the text is not to give pride of place to belief as a justifying state, but rather to stress that belief is the sort of thing that both stands in need of justification and justifies.

The truth of explanations in terms of our schema, therefore, is highly sensitive to whether the subject of the explanation is prepared to endorse the explanation herself. But it is a necessary condition of possessing a capacity to judge that enough of one's performances are explainable by means of our schema. Thus, it follows that possessing a capacity to judge requires having some degree of sensitivity to one's own beliefs and one's reasons for them — in particular, a degree high enough to allow some such explanations to be true.

This gives us a way to formulate what the system described a few paragraphs back lacks. Correct explanations of that system's transitions between states involve no dependence on a special kind of attitude to what goes in and out of its belief-box (the system doesn't even have any provisions for such an attitude). Thus, they are not subject to explanations of the appropriate form. Even if we added a special set of second-order registers that reliably tracked the changes in the contents of its first-order registers, the explanations of those first-order changes would *still* not be sensitive to the system's second-order contents. This entails that the system's transitions between first-order states are not made for reasons, even though (as judged from the outside) they may be as rational as we like. But this entails that our system lacks a capacity to judge.

4. Even though the epistemic life of ordinarily mature and reflective subjects involves a significant amount of higher-order belief, as I have been insisting the reflexivity of belief as such is not a capacity for second-order belief. The "taking oneself to believe" that is involved in judging is not a second-order judgment or belief, and it does not necessarily issue in one. It is a form of awareness more basic than that, which forms part of the background against which second-order belief is to be understood.

A subject may be able to exercise her capacity to judge by treating her existing beliefs as giving her reasons for new judgments, and her new judgments as giving her reasons either to retain or to revise old beliefs, without thinking of herself as a subject of beliefs. This seems clear, even if

we consider the above case. What was important about Fred's thinking was that it was subject to a certain kind of explanation, which Fred himself could endorse. But explanations of this kind do not assign an explanatory role to second-order beliefs as such. Of course, *believing* the explanation would itself be a second-order state; but the reflexivity of judgment does not require one to *actually believe* true explanations of the appropriate form, but only to be *disposed* to do so. This disposition is grounded, as I have been suggesting, in a feature of first-order belief itself.

Thus full-blown self-knowledge is not guaranteed by the reflexivity of belief alone, since the latter does not need to involve any beliefs about one's own mental states, and (full-blown) knowledge entails belief. But the reflexivity of belief offers the necessary background against which a capacity for second-order judgment, and thus full-blown self-knowledge, can be understood. Similarly, as I will argue in the next section, the reflexivity of belief by itself does not suffice for full-blown a priori knowledge; but it provides a piece of the necessary background against which a priori knowledge must be understood.²⁷

²⁷ The question naturally arises whether one needs to be at least *capable of* such second-order judgment and belief in order to be capable of reflexive first-order beliefs. In other words, does one need to have something like our concepts of belief, judgment, and reason in order to have reflexive first-order beliefs?

On the one hand, it seems a bit disingenuous to insist that a creature with a fully reflexive first-order cognitive faculty — a creature that is aware of its own first-order beliefs and its reasons for them — might nevertheless *lack* the concepts of belief or reason. What more should we require as evidence that it possesses them? It seems odd to say that our creature is capable of being aware of itself *as* believing for reasons, but nevertheless it lacks the concepts of judgment and reason. The distinction between having a capacity to be aware of something *as* an *F* and possessing the *concept* of an *F* seems obscure. On the other hand, we can find intuitively plausible examples of creatures that have a capacity to judge, but do not seem to have a capacity for second-order thought. Arguably, something like this might actually be the case with young children or non-human animals. Unless we take these cases to bring the point about the reflexivity of belief itself in question, shouldn't we take them to prove that we must make room for pre-conceptual capacities for self-awareness?

I will not attempt to resolve this issue here. Let me note, however, that I do not think that it is altogether implausible to take the capacity to judge — at least *our* capacity to judge — to be an acquired rather than an innate skill, so that subjects count as full-blown judges only at the end of a developmental process. This would entail that we do not share the same capacity to judge with creatures that lack the necessary reflective concepts. But it need not entail that we do not share something more generic with them: perhaps the capacity to have an intelligent, and adaptive, take on the world around us. (See McDowell (1994: Lecture VI).)

5. As I suggested early on in this chapter, the account of self-awareness just given combines with the holistic account of concept-possession to show that a judging subject must not only be sufficiently reliable in conforming to the norms of her practice; she must, moreover, be sufficiently reliable in *recognizing* particular instances of those norms. This, I take it, is the fundamental capacity on which our full-blown capacity for a priori knowledge is based.

According to what I have been arguing so far, a mature subject's thought is reflexive, in the following sense: if our subject believes p for reason R , then she normally is aware — even if she does not *judge* — that she believes p for reason R , and thus — by the same token — she takes R to *be* (objectively) reason for p .

Now, of course, on occasion the subject's reasons for a certain belief will be bad; and in that case the normative judgment she is disposed to make — namely, that R objectively is reason for p — will be false, and thus not a candidate for knowledge at all. Moreover, in some of the cases where R really *is* a reason for p , this will be so only contingent on further, empirical background knowledge. For example, the fact that Fred's belief that Mary is sick with a fever gives him good reason to believe that she will not be at the meeting depends on Fred's background knowledge — for example, his knowledge that normally one has to be reasonably healthy to participate in the meeting, that being present at the meeting is not of life-and-death importance to Mary, and so on. Thus, although Fred knows that his belief that Mary is sick gives him good reason to believe that she will not be at the meeting, his knowledge is empirical, not a priori.

On the other hand, given the holistic constraint of concept possession, it must be the case that, for a sufficiently broad and central range of the norms constitutive of her concepts, the subject is reliably disposed to conform to them. But, now, the account of self-awareness we have guarantees that this is no longer a mere claim of reliability: dispositions to judge or infer are no longer blind. Suppose that our subject is in fact reliably disposed to infer in accordance with the rule

of double negation. This, on the present account, entails that, on the occasions where she does infer in accordance with it, she is aware that she believes that p for the reason that $\sim\sim p$; thus she takes it that $\sim\sim p$ is a good reason to believe that p .²⁸ But, given this context, there is a perfectly intelligible sense in which this awareness is a case of our subject's *recognizing* the correctness of an instance of the law of double negation. In a perfectly intelligible sense, therefore, our subject now can be said to infer in accordance with the law of double negation *because she recognizes* the correctness of its instances. If she explicitly goes on to judge such an inference to be correct on this basis, this is a case of knowledge. This, as I have been suggesting throughout, is our fundamental capacity for a priori knowledge.²⁹

6. The question I have been trying to answer in this chapter was how to account for a subject's justification in conforming her thinking to norms which, as a matter of fact, are constitutive of her concepts. The answer I gave focuses on the self-awareness characteristic of conceptual thought: a subject is, at least in a broad and central enough range of cases, justified in conforming her thinking to those norms because she recognizes the correctness of their instances.

²⁸ As noted earlier, there is a sense in which this inference too depends on — possibly empirical — background knowledge: the subject's reasons for $\sim\sim p$ must be stronger than whatever reasons she may have to reject p . But the case is not at all like the empirical case discussed in the text: the rational connection between $\sim\sim p$ and p , and Fred's recognition of it, do not *themselves* depend on background knowledge, as the connection between Fred's belief that Mary is sick and his belief that she will not be at the meeting did. It is only what Fred is entitled to do with his recognition that depends on empirical knowledge.

²⁹ Would our subject also be in a position to know that she knows that p follows from $\sim\sim p$? Normally, I think she would. Consider what the situation would be like if she did not. Our subject knows that p follows from $\sim\sim p$. Presumably she also knows that she judges p to follow from $\sim\sim p$. If this is so, then what is there to stop her from arguing "I judge p to follow from $\sim\sim p$; and p follows from $\sim\sim p$; therefore, I don't mistakenly judge that p follows from $\sim\sim p$?" But it seems intuitive that our subject cannot use such an argument to reassure herself about her a priori judgments. (A similar point is widely, though not universally, acknowledged in the case of perception. (See Cohen (2002), Davies (2004), Pryor (2000; 2004) and Wright (2004) for different takes on the issue.) I suggest that the problem with such arguments is precisely that, normally, if we attribute to a subject a priori knowledge that p follows from $\sim\sim p$, we also — and thereby — take her to know (or to be in a position to know) that she knows this fact. Given that, on the present view, basic a priori knowledge is attributed on the basis of a self-conscious capacity to judge, such self-knowledge should not be surprising.

Now, of course, the overall aim of this dissertation has been to give an account of a priori knowledge in general, and the interest in this question was generated only because, intuitively, it appeared that the capacity to conform one's thinking to the norms constitutive of one's concepts is somehow fundamental to a priori knowledge in general. Indeed, I think that it is. I will close this chapter by revisiting the question of how to get full-blown a priori knowledge out of exercises of this capacity.

Recall our example of a subject's coming to know that the law of double negation is a valid rule of inference (assuming, of course, that classical logic is correct, and therefore that this is something we can know). Intuitively, it seems that if anything interesting is knowable a priori, basic logical laws are. As I suggested earlier, it is intuitively plausible that at the basis of any account of how one comes to know that a basic law of logic is valid there must be an account of how one recognizes particular instances of the law. In one way or another, one's knowledge that the law of double negation is valid depends on one's being able to recognize the validity of particular inferences in accordance with that law.

As I have been trying to argue, this latter ability is to be explained in terms of the reflexivity of belief. In the first instance, of course, the reflexivity of belief explains how, when we actually *do* infer in accordance with — for example — double negation, we are justified in doing so: we are justified because we recognize the inference as a good one. But the same ability can plausibly be exercised “offline”, when one is not in the business of actually *making* inferences, but rather simply in the business of evaluating them (perhaps one is doing a piece of logic homework). Thus the same sort of capacity that justifies the subject in making an inference in accordance with the law of double negation can be exercised when the subject merely entertains, for the purposes of evaluation, the same inference.

There is, of course, more to be said about how one moves from the ability to evaluate particular inferences to knowledge that the law of double negation is a valid inference rule. I will not go into this question here in any detail. But it seems plausible that, if one is capable of correctly evaluating particular instances of the law, then one may also — given sufficient background knowledge — figure out that the inferences work because of the logical form of the premise and conclusion. Supposing our subject to be capable of explicitly theorizing about the validity of her inferences, it is plausible that she is in a position to recognize that in any particular case the validity of an inference in accordance to the law of double negation is due to the distinctive contribution of negation to the relevant propositions, rather than to any particular feature of the case: thus she can tell that, no matter what proposition p is, if one believes that $\sim\sim p$, then one also ought to believe that p . But then, she is now in a position to know that the law of double negation is a valid rule of inference.

Appendix A
ON THE INDIVIDUATION OF FREGEAN THOUGHTS

1. INTRODUCTION

Throughout this dissertation I rely on a Fregean conception of thoughts. I introduced this conception in Chapter 1, by contrasting it to Russell's conception of propositions. Although both Russellian propositions and Fregean thoughts are conceived of as the objects of intentional mental states and acts, the availability to the mind of Russellian propositions is, in a sense, merely an accident: nothing about their nature follows from the nature of thinking. The nature of Fregean thoughts, by contrast, is not similarly independent of the nature of thinking: Fregean thoughts are individuated by reference to the same principles of reason that govern our thinking. This is not just an epistemological claim, concerning the criteria we apply in practice to attribute content to people; it is a metaphysical claim, concerning what Fregean thoughts are. In this Appendix, I will elaborate a bit further on the question of the individuation of Fregean thoughts.

2. A LINGUISTIC APPROACH

Frege's notion of thought is a descendant of his earlier notion of "conceptual content" (*begrifflichen Inhalt*).¹ As such, it is meant to capture, from the perspective of the subject of a mental state or act, that feature or aspect of it that is relevant to its truth or falsity. In thinking a thought we represent

¹ Frege (1967: 6).

things as being a certain way, and in judging it we take them to be this way. Our mental states or acts, of course, usually also have features which are not relevant to how they represent things as being; these features, according to Frege, do not belong to the conceptual content of the mental states or acts — they are not features of the relevant thoughts.

Crucially, thoughts are the sort of thing that can be communicated in language. This means neither that all thoughts must be capable of being communicated in this way, nor that natural language provides an ideal vehicle for just any thought. Nevertheless, in a core of central cases, sentences in natural language not only express thoughts, but also exhibit a syntactic structure that corresponds to the structure of the thoughts themselves. The thought in these central cases is the *sense* of the sentence, and the constituent parts of the thought are assigned as senses to the corresponding constituent parts of the sentence. The sense of a linguistic expression is, in general, the “mode of presentation” of its referent or “meaning” (*Bedeutung*) — that is, the way the referent is thought of by one who understands the expression.² On Frege’s official exposition the referent of a singular term is an object, while that of a predicative expression is a “concept” — a function, that is, from the set of objects (or n-tuples of objects) to the set of truth-values. Sentences are conceived of as complex singular terms, with truth-values as their referents.³

The connection between language and thought makes it natural to look at language for clues concerning the individuation of thoughts. More specifically, one might hope for a principle for the

² It is sometimes assumed that the Fregean sense of names at least must be something like a cluster of definite descriptions which are known (or at least believed) to hold uniquely of its referent. If that were so, then the question of the individuation of at least these senses would be quite easy to answer. But I think there is no compelling reason to take the sense of names to consist in such descriptions.

In fact, given Frege’s equation of *grasping the sense* of a referring expression with *understanding* that expression, and the compelling arguments to the effect that understanding a name does not consist in coming to know — or at least believe — a definite description, it seems that we have good reason to avoid attributing to Frege that view.

³ Frege (1997b; 1997e).

individuation of linguistic senses, in the hope that this will take us some way towards a principle for the individuation of thoughts in general.

It is possible (although not mandatory) to extract from Frege's essay 'On Sense and Reference' the following simple proposal for the individuation of linguistic senses, which I will call the Intersubstitutability Criterion: two expressions in a language *L* have the same sense only if they are intersubstitutable without change in truth value in all contexts.⁴

As Frege's discussion makes clear, the relevant contexts for present purposes are the 'that'-clauses of indirect speech attributions of propositional attitudes. Frege presents this proposal as an application of Leibniz's Law, which — in the formulation that is of relevance here, at least — states that expressions have the same *reference* only if they can be substituted for each other in every context. At first sight, the contexts generated by the 'that'-clauses of indirect attributions of propositional attitudes seem to generate counterexamples to this principle. We are inclined to accept that:

(1) Jones believes that Phosphorus is a planet

may differ in truth-value from:

(2) Jones believes that Hesperus is a planet

even though, as we (unlike Jones) know, 'Phosphorus' and 'Hesperus' both refer to the planet

Venus. According to Frege, however, such pairs of sentences are not counterexamples to Leibniz's

Law, for expressions used in the 'that'-clauses of indirect attributions do not refer to the referents

they have in ordinary contexts, but rather to their *senses*. Taking 'believes' in this context to express a

⁴ The Intersubstitutability Criterion as formulated gives us a necessary condition for identity, but not a sufficient one. In fact, no equally simple sufficient condition of the identity of sense can be formulated in the same spirit. This is because, prior to an exhaustive examination of the expressive resources of *L*, we have no guarantee that for every two expressions with distinct senses there will be a sentence of *L* such that substituting one expression for the other in it will result in a change in truth-value. The Intersubstitutability Criterion, therefore, can be strengthened to a biconditional only on the assumption that we are dealing with a rich enough language. Since the points I want to make concerning the Intersubstitutability Criterion do not hinge on having a sufficient condition for identity, I will ignore this complication in what follows.

two-place relation between a subject and a thought, the conclusion Frege draws from the possibility that (1) and (2) may diverge in truth-value is that the *sense* — the *thought* — expressed by “Phosphorus is a planet” differs from that expressed by “Hesperus is a planet”. Given that we take the sense of both of these sentences to be uniquely determined by the senses of their constituent expressions, it follows that this difference must be attributed to a difference in sense between ‘Hesperus’ and ‘Phosphorus’, since all their other constituents are common to both sentences. If we generalize this intuitive line of thought, the result is the Intersubstitutability Criterion.

It is not, as I already suggested, entirely clear whether Frege ever subscribed to the Intersubstitutability Criterion. In fact, I think it is easy to see that the principle as stated above is not satisfactory. The intuitive idea motivating the Intersubstitutability Criterion is that thoughts give us a subject’s perspective on the world, and therefore that the principles of individuation of thoughts should be sensitive to the way in which we attribute intentional mental states to each other. Even though there is something clearly right about this, it is not adequately captured by the Intersubstitutability Criterion. This, I believe, is because this criterion is too closely tied to the way linguistic expressions of propositional attitudes function. But there is no guarantee that linguistic usage is systematic enough to provide an adequate ground for a theoretical account of the individuation of thoughts.

We can see this by considering some counterexamples that have been proposed in the literature, and the ways in which the Intersubstitutability Criterion might be defended in the face of them. I hope it will become clear that the criterion is not viable, as it requires us to make distinctions between senses that, intuitively, are not well-motivated. This, I take it, constitutes evidence that the individuation of thoughts cannot be simply read off from the behavior of propositional attitude expressions in natural language.

Consider the case of an English-speaking subject who does not know that the words ‘furze’ and ‘gorse’ are different names for the same plant (the example is discussed by Kripke⁵). Since it is common knowledge that furze and gorse are the same plant, the case is significantly different from Frege’s case of ‘Phosphorus’ and ‘Hesperus’.⁶ ‘furze’ and ‘gorse’ are intuitively synonyms in English, whereas there is no similar pressure to take ‘Phosphorus’ and ‘Hesperus’ as synonymous — certainly not prior to the relevant astronomical discovery. If Jones is not party to this common bit of botanical knowledge, however, it would appear possible that the sentences:

(3) Jones believes that furze is spiny

(4) Jones believes that gorse is spiny

may differ in truth-value. According to the Intersubstitutability Criterion, this should lead us to conclude that ‘furze’ and ‘gorse’ differ in sense.

The defender of the Intersubstitutability Criterion, therefore, would have to insist that, even though ‘furze’ and ‘gorse’ intuitively express the same sense for most people, nevertheless they express different senses in the mouth of our ignorant subject. Moreover, she should add, when we evaluate a sentence which includes an intentional verb governing a ‘that’-clause, we should be careful to take the expressions within the ‘that’-clause to refer not to their ordinary senses, but to the senses attached to them by the person denoted by the grammatical subject of the verb. Of course it may turn out that the two senses are identical; but we cannot assume this at the outset.

This move is effective against our puzzle, but it is not clear that the defender of Fregean sense should be happy with it.⁷ Sense is supposed to be an objective and, for the most part, shared

⁵ Kripke (1979).

⁶ Frege actual examples are ‘the Morning Star’ and ‘the Evening Star’; but since these expressions look, grammatically, like definite descriptions, it may be better to use the unstructured expressions I use in the text.

⁷ Kripke briefly considers such moves in response to his puzzle, but he quickly dismisses them as implausible. I think the reason he dismisses them, however, is just that he assumes that sense must be something like a cluster of descriptions associated with a linguistic expression. But, as I suggested earlier, there seems to be no compelling reason for that assumption.

aspect of language. Of course Frege did envisage co-referential names that differ in sense, and therefore we should certainly expect our principle of individuation to allow for such cases. We should, in other words, expect no guarantee that ‘furze’ and ‘gorse’ will *never* differ in sense, in the mouth of any English-speaking subject who uses both to refer to the same plant, no matter what their individual histories and the state of their botanical knowledge. The problem is not that the Intersubstitutability Criterion demands making such distinctions at the level of idiolects; it is, rather, that it demands making them far too easily, regardless of any intuitively relevant contextual details.

As we already saw, in any case where (3) and (4) intuitively differ in truth value the Intersubstitutability Criterion will classify ‘furze’ and ‘gorse’ as Jones uses them as differing in sense. But further details of the case might make this hard to accept. Jones, for example, might have picked up the word ‘furze’ only recently, just from someone else’s conversation concerning a plant that grows in dry areas. Intuitively, we would say that she has only a limited understanding of its sense. If she is reasonable, she will presumably agree with our judgment, and therefore we may suppose that in her use of ‘furze’ she will intend to defer to people who know more about it (perhaps to the people she picked the word up from). Given this description of the case, it may be appropriate to judge that (3) is false while (4) is true: Jones herself, for instance, would in all probability refuse to assent to (3), given how shaky her understanding of ‘furze’ is. If this is right, the Intersubstitutability Criterion would require us to classify ‘furze’ and ‘gorse’ as differing in sense for Jones. Intuitively, however, the word ‘furze’ in Jones’s mouth has just the same sense as in the mouth of the people she intends to defer to; and, by hypothesis, *that* sense is not different from the sense of ‘gorse’, as both Jones and the more knowledgeable subjects use it.

Moreover, the proposed maneuver is not going to work in general, for there are examples involving embedded attitude ascriptions, which do not rely on the subject’s ignorance of the relevant synonymies. Consider, for example, Smith, who is a botanical sophisticate — at least to the extent

of being party to the knowledge that furze and gorse are the same plant. It is thus implausible that the words ‘furze’ and ‘gorse’ will differ in sense as she uses them. Nevertheless, it is still plausible that the following will differ in truth-value:

(5) Smith believes that everyone believes that all furze is furze

(6) Smith believes that everyone believes that all furze is gorse

If Smith is minimally reflective, (5) will plausibly be true and (6) will be false, since she will realize that not everyone is as competent in her use of the words ‘furze’ and ‘gorse’ as she is — she may know that her friend Jones, for example, is not. But if this is right, then by the Intersubstitutability Criterion ‘furze’ and ‘gorse’ should again differ in sense in Smith’s mouth. Since our assumption was that they don’t, this tells against the criterion itself.⁸

3. AN ALTERNATIVE APPROACH

1. The source of trouble here, I believe, is that we have rested too much weight on linguistic considerations concerning the way propositional attitude attributions work in English. It would be convenient if a simple formal criterion were forthcoming through such considerations. But the examples we have considered so far seem to suggest that the linguistic practice of psychological attribution does not by itself permit the formulation of such a criterion.

⁸ This is a version of Mates’s puzzle, discussed in Burge (1978). Frege’s view of an hierarchy of senses significantly complicates the case. On a view that embraces the hierarchy, the point must be that the *second*-level sense of ‘furze’ must differ from the *second*-level sense of ‘gorse’ in Smith’s mouth. Since higher-order sense is determined by one’s take on another’s thinking, this is not nearly as implausible as taking Smith’s *first*-level senses to differ. Still, such cases show that combining the Intersubstitutability Criterion with Frege’s hierarchy of senses is pretty much guaranteed to have the result that no two words can have the same second (or higher) level sense, for any minimally reflective subject. It is not clear that a proponent of the hierarchy should be comfortable with this.

If this suggestion is right, then we evidently need a different approach to the individuation of thoughts. The approach I am going to suggest is motivated, just as the Intersubstitutability Criterion itself was, by the idea that the identity of Fregean thoughts is fixed by their place in our rational economy. We can spell this out somewhat by observing that an account of the individuation of thoughts should combine with an account of the propositional attitudes to give us the resources to understand people as rational thinkers.⁹ On the present proposal, however, it is not assumed that the way propositional attitude attributions work in English (or any other natural language) is always a perfect guide to the *nature* of these attitudes (or, at least, to the most interesting or central among them). Instead, we need to develop our account by specifying explicit, intuitive constraints on the sorts of attitudes that are central to the individuation of thoughts.

We can begin by considering the following principle suggested by Evans in his interpretation of Frege (the formulation is adapted in order to eliminate reference to sentences):

The Intuitive Criterion of Difference: A thought *T* must be different from the thought *T'* if it is possible for someone to understand both at a given time, while coherently taking mutually exclusive attitudes towards them, [e.g.] accepting one while rejecting, or being agnostic about, the other.¹⁰

It is clear that the criterion as stated here is quite rough. For one thing, it does not in general apply to thoughts expressed by sentences containing indexical expressions. For example, Frege famously held that each person has a unique way of referring to herself, expressible by the first-person pronoun ‘I’. It follows that two sentences uttered by two different individuals, both of which involve the use of ‘I’, must always express different thoughts. Yet it is not possible for anyone to rationally hold contradictory attitudes to these two thoughts, not because doing so would be irrational, but rather because there is no one person who can hold *any* attitude to both of these

⁹ See McDowell (1998b).

¹⁰ See Evans (1982: 18-9). This principle is explicitly appealed to by Frege in a number of places. See, for example, Frege (1997b: 156; 1997c: 332-333; d: 128).

thoughts.¹¹ I will not here try to formulate further criteria to take care of these cases, since the points I want to make can be made without reference to them.

Evans's criterion, as well as the proposal I will go on to make in a similar spirit, are attempts to spell out the idea that thoughts are individuated in terms of their role in the mental economy of rational subjects. They depend, therefore, on psychological notions, and primarily on the propositional attitudes. The crucial difference between this approach and the one that led to the Intersubstitutability Criterion is that the present approach is much less dependent on linguistic considerations: the relevant propositional attitudes are specified by means of explicit constraints, thus minimizing reliance on the vastly complicated conventions governing the use of verbs like 'to believe'. Giving a full account of how expressions for propositional attitudes work in a natural language is a difficult problem; the present approach aims to sidestep it as much as possible.

What, on the contrary, will need to be addressed in later sections is the nature of the constraints imposed on the attitudes relevant to the individuation of thoughts. The crucial notions used in the Intuitive Criterion of Difference are those of understanding and — in effect — of the rational incompatibility between attitudes. These are notions that need to be explicitly discussed. Much of this Appendix is devoted to this task. But before getting to the discussion of these notions, we need to examine some more basic reasons why we cannot be fully satisfied with the Intuitive Criterion of Difference.

2. The Intuitive Criterion of Difference as formulated above gives us only a sufficient condition of difference (or, alternatively, a necessary condition of identity) between thoughts. Could we

¹¹ The problem here is not formal, since the Intuitive Criterion of Difference as formulated does not purport to give a *necessary* criterion of difference, only a sufficient one. The problem, rather, is that such first-personal indexical thoughts are altogether outside the reach of the Intuitive Criterion, and therefore something further has to be said about their principles of individuation.

strengthen it to a biconditional, so as to get a necessary condition as well? We could not. For consider the thoughts expressed by the following two sentences:

(7) Jones is Jones

(8) Smith is Smith

It is intuitively clear that these are two different thoughts, yet it is also clear — even without a theoretical account of understanding or rational incompatibility — that no one could understand both while rationally taking incompatible attitudes towards them.¹² One could rationally doubt either (7) or (8) only by questioning the law of identity (I am assuming that there is such a thing as rationally doubting a fundamental logical law; if this is thought to be problematic, the argument in the next paragraph would still apply). But if a subject questions the law of identity in one of the two cases, it would obviously be irrational for her not to question it in the other. Thus, for a rational subject the thoughts expressed by (7) and (8) stand or fall together, even though they are — intuitively — distinct.

A parallel problem would arise for the biconditional version of the Intuitive Criterion of Difference in the case of thoughts which cannot be comprehendingly endorsed (or rejected), on pain of irrationality. If we strengthen the Intuitive Criterion of Difference to a biconditional, then that will have the result that there can be only one thought such that, if one understands it, one rationally has to endorse it, and only one thought such that if one understands it one rationally has to reject it. It is clear that if one wanted to defend the existence of thoughts such that understanding them rationally requires endorsement (or rejection), then one could not accept this consequence. It is, of course, controversial whether there are *any* thoughts such that understanding them is incompatible

¹² The possible failure of either Jones or Smith to exist might complicate the picture, at least on some views of non-referring names. But if we stick with Evans's view, if either 'Jones' or 'Smith' fail to refer, the corresponding sentence expresses no thought, and therefore there is no such thing as *understanding* it. Thus the formulation in the text is correct, even if questions of existence are allowed to bear on the case.

with either rejecting or endorsing them; but this should not be ruled out by merely formal considerations concerning the individuation of thoughts.

Returning to (7) and (8), it is easy to see why the biconditional version of the Intuitive Criterion of Difference fails. The Intuitive Criterion of Difference is not explicitly sensitive to the internal structure of thoughts; yet it is clear that what differentiates the thought expressed by (7) from the thought expressed by (8) is that in (8) ‘Smith’ has been substituted for ‘Jones’.

It is, in any case, independently intuitive that thoughts should be treated as functionally composite, in the sense that they are uniquely determined by their parts and the ways in which these parts are combined. This gives us a necessary condition of difference between thoughts (and, conversely, a sufficient condition of identity):

Compositional Criterion: If two thoughts are distinct, then they must differ either in their components or in their mode of composition.¹³

Now, of course, this principle cannot take us very far, unless we can say something about the individuation of thought-components and modes of composition. From a Fregean perspective, however, this is not such a straightforward matter. Supposing that we had access to the internal structures of thoughts, as well as criteria of identity for thought-components and modes of composition, then we could of course move back up and formulate criteria for the identity of thoughts in terms of their composition from those more basic elements. Given Frege’s “context principle”, however, this direct route is not available to us: as I already stressed in Chapter 1,

¹³ The converse does not straightforwardly hold, since — as Frege repeatedly emphasized — thoughts can be decomposed in many different ways: the thought ‘*Fa*’, for example, can be seen to consist of the first-level concept expressed by ‘*Fξ*’ predicated of *a*, or, alternatively, of the second-level concept expressed by ‘*Φa*’ predicated of the first-level concept expressed by ‘*Fξ*’. To formulate a condition both sufficient *and* necessary for the identity of thoughts, therefore, we need the set of all possible decompositions of a given thought.

according to Frege it is only through their contribution to whole thoughts that thought-components and modes of combination are individuated.¹⁴

Even independently of Frege's letter, it is in any case hard to see how an appeal to independent criteria for the individuation of thought-components and modes of composition could be warranted at this point. Thoughts are the objects of our propositional attitudes — they are the things we judge, believe, question, anticipate, and so on. Thought-components and modes of composition, by contrast, are available to us only in virtue of second-order thinking about thoughts themselves.

As has been noted by many commentators on Frege, we get access to the internal structure of thoughts not by narrowing our focus, as it were, so as to look at the interior of thoughts, but rather by expanding our view to encompass a whole system of thoughts and attitudes.¹⁵ To determine the internal structure of a thought we need to see that thought in rational relations — for example, relations of logical entailment or inconsistency — with other thoughts.

To see how this would work in the context of the present proposal, consider first a simple example along the lines of those discussed in the previous section. Consider the following sentences:

(9) Hesperus is Hesperus

(10) Hesperus is Phosphorus

How are we to determine whether these express the same or different thoughts in the mouth of an astronomically ignorant subject, Jones? Well, if we have evidence that our apparently rational and generally competent subject is disposed to assent to (for instance) 'Hesperus is a planet', but shows no disposition to infer from this to the sentence 'Phosphorus is a planet', then we have some

¹⁴ The principle is first formulated in Frege (1980) that is, before Frege develops his conception of sense and reference. It is natural, however, to take Frege to have held the analogous thesis for thoughts and their components.

¹⁵ For a very useful discussion of the topic, see Ricketts (forthcoming).

pressure to take ‘Hesperus’ and ‘Phosphorus’ to express different senses in her mouth, and hence some pressure to take (9) and (10) to express different thoughts. Of course this pressure may be counteracted — if, for example, Jones is a logician notorious for her rejection of Leibniz’s Law, then we may have reason to think otherwise. But the point is not to decide this particular case: the point of the present example is just to show how taking into account a system of propositional attitudes, rather than considering them one by one, gives us access to the *components* of thoughts, and thereby helps fix the individuation of whole thoughts.

Once again we see that it is necessary to appeal to the same idea of the rational combinability of attitudes that was central to Evans’s Intuitive Criterion of Difference. Some of the rational combinations of attitudes and thoughts we need to consider will be governed by formal deductive principles, like Leibniz’s Law in the previous example. Others, although still relying on deductively valid inferences, may not be possible to capture in a formal system. Thus the truths of arithmetic may be plausibly constitutive of our thoughts involving numbers, but they cannot all be proved in a consistent formal system.

Also relevant may be relations which are neither deductive, nor can be plausibly traced back to one or another particular concept, or even to relations between a fixed set of concepts. It is often remarked that a constraint upon interpretation is that we should, as far as possible, interpret a subject’s thinking by attributing to her concepts that capture “natural kinds” in her environment.¹⁶ Normally, we would interpret a subject as thinking emeralds to be green rather than grue — or, rather, a 21st century analogue of grue — for example, precisely because the former predicate is suited to our inductive practices. If so, then principles of inductive inference are constitutive of a hugely significant — though open-ended — class of concepts.¹⁷

¹⁶ See Lewis’s “eligibility constraint”, in Lewis (1999).

¹⁷ It is worth pointing out that the approach to the individuation of thoughts I have outlined here is compatible with truth-conditional approaches in semantics. The idea of this compatibility was (arguably)

4. UNDERSTANDING AND PROPOSITIONAL ATTITUDES

1. As I pointed out earlier, the move away from the details of the linguistic practice of propositional attitude attribution has a point only if combined with explicit theoretical constraints on the propositional attitudes relevant to the individuation of thoughts. This enables us to cut through the enormously complex system of conventions governing the use of propositional attitude expressions in natural language. In this section, as well as the ones that follow, I will be considering the nature of such constraints.

The examples above all turned on intuitive judgments about interpretation. But it should be emphasized that the thesis about individuation I am proposing is not a verificationist one, to the effect that thoughts are individuated in terms of the judgments we — or a suitably idealized version of ourselves — would make if placed in the position of the interpreter in the corresponding scenario. The thesis concerns which combinations of attitudes and thoughts are rationally available to a subject, regardless of how — or even whether — anyone recognizes them to be so. Of course what a particular subject is disposed to say or do constitutes part of the evidence we rely on in identifying her thoughts. But this evidence is supposed to concern an independent subject-matter,

already present in Frege, and has been developed more recently by McDowell and Evans in the framework of the Davidsonian project of semantics for natural languages (Evans (1982: ch. 1); McDowell (1998b)). In a Davidsonian theory of meaning, material biconditionals specifying the truth-conditions of sentences in the target language are supposed to be derivable as theorems in a recursive theory of truth, whose axioms specify the reference of the simple expressions of the language and whose recursion clauses specify its primitive modes of combination. But what qualifies a given recursive theory of the right form as, first, specifically a theory of *truth* for the target language, and, second, as the one among the many extensionally equivalent theories of truth that captures the speakers' intended meanings? According to McDowell and Evans's approach, the answer must be that the thought that the sentence on the left-hand side of a T-theorem would express in the mouth of the native speakers is the same as the thought expressed in the meta-language by the sentence used on the right-hand side. Such an approach to truth-conditional semantics evidently requires an account of the individuation of thoughts (as well as an appeal to a pragmatic theory for utterances in the target language, which I am here ignoring).

which is neither exhausted by the totality of evidence we can possibly have about it, nor constituted by the judgments we make based on that evidence.

This serves to deflect a kind of Quinean objection that might surface here.¹⁸ I have suggested above that certain inferential patterns are constitutive of our concepts. It is, for example, constitutive of the concept *green* that the attitude of believing “*x* is green” is incompatible with the attitude of rejecting “*x* is colored”. A Quinean might object to this view by challenging its proponents to give an account of *which* patterns are constitutive of the concept. As the Quinean would point out, we may find that our subject’s inclination to assent to the sentence “*x* is Fred’s favorite color” conditional on her assent to “*x* is green” is just as strong as her inclination to assent to “*x* is colored”, conditional on her assent to “*x* is green”. Assuming that dispositions to assent to or dissent from sentences are all we have to go on when we attempt to identify our subject’s thoughts, the Quinean would then deny that we have any reason take the latter pattern to be privileged over the former.

From the present perspective, however, this challenge may simply be dismissed. The former pattern is not constitutive of the concept *green*, simply because it is not true that believing that *x* is green stands in any interesting relation to not believing that *x* is Fred’s favorite color. A relation between the two could only hold because of some further attitude — for example, our subject’s believing that green is Fred’s favorite color. The pattern involving *all three* of these attitudes might have something interesting to teach us about some of the concepts involved, but there simply is no interesting pattern involving just the first two. The fact that no amount of observation of the subject’s inclinations to make the relevant transitions suffices to distinguish them in status is simply irrelevant here.

¹⁸ The style of objection I am imagining follows Quine (1980).

Of course, in giving this response I am making use of my understanding of the term ‘green’; thus I am not respecting the Quinean’s restriction of admissible evidence to the subject’s dispositions to assent or dissent. But even in a radical case, where we — as interpreters — do not yet understand our subject’s words, what matters for the identity of their thoughts is not their dispositions to assent to or dissent from sentences, but rather which combinations of attitudes are available to them. Perhaps our only hope to get to those is through close observation of their dispositions to assent to and dissent from sentences which we do not as yet understand. Perhaps, too, this will enable us to make only the crudest distinctions among the thoughts expressed by these sentences. But this, from the present perspective, says nothing at all about the rational relations among our subjects’ attitudes, or — accordingly — about the identity of their thoughts.

2. Thus the notion we need to consider is that of the rational availability of certain combinations of attitudes to a subject. This is the notion that, as we saw, is central to the proposal developed in §3. We can begin by considering the question whether one could rationally take contrasting attitudes towards the thoughts expressed by the following sentences, while understanding both of them:¹⁹

(12) Furze is spiny

(13) Gorse is spiny

It seems that one could believe one of them while being unsure of, or even rejecting, the other. All that is required for this to be the case is a bit of botanical ignorance: one may, for instance, not know that furze is the same plant as gorse, and be unsure whether furze is spiny or not. We would normally take such ignorance to impugn neither one’s rationality nor one’s understanding of the

¹⁹ I am assuming that we understand which attitudes are, intuitively, “contrasting” in the relevant sense: for example, endorsing and rejecting, or being agnostic, are contrasting attitudes in the present sense. This does not mean (as will become evident in the ensuing discussion) that it is impossible for a subject to hold such attitudes in respect to the same thought — even without irrationality. I think that we can help ourselves to an intuitive understanding of this notion: if one knows what belief is, then one normally also knows that it contrasts (in our sense) with *dis*belief.

relevant terms. If we accept these intuitive judgments, then it should follow that (12) and (13) express different thoughts, at least in the mouth of the botanically ignorant subject. This is a direct consequence of Evans's Intuitive Criterion of Difference. Yet, to the extent that we take this ignorance not to impugn our subject's understanding of either 'furze' or 'gorse', (12) and (13) do not express different thoughts.

The difficulty is that, although (12) and (13) express the same thought, a subject who understands both sentences need not know that they do. Insofar as this is possible, our criteria are not guaranteed to give the right results, for if the subject does not know that it is the same thought she is considering twice over, then it may not be a failure of rationality on her part to take contrasting attitudes towards it. Thus, if we are to employ the notion of the rational availability of combinations of attitudes to a subject in determining questions about the identity of thoughts, we must rule out the possibility that the subject may fail to recognize that two distinct episodes of conscious thinking involve the same thought, through no failure of rationality or understanding.

It is frequently assumed that Fregean sense is *transparent*, in the following sense: if one grasps a sense then, on every occasion when it is occurrently present in one's consciousness, one must be in a position to recognize it as the same (at least barring failures of memory, attention, and the like).²⁰ There is evidence, however, that Frege *rejects* the transparency of sense. As his account of the role of definition and logical analysis in mathematics shows, recognizing a correct analysis is often a significant cognitive achievement:

But how does one judge whether a logical analysis is correct? We cannot prove it to be so. The most one can be certain of is that as far as the form of words goes we have the same sentence after the analysis as before. But that the thought itself remains the same is problematic.²¹

²⁰ See, for example, Dummett (1975: 131); Levine (1998:433). A notable exception is Burge (1990). I think this view is in part responsible for the common assumption that a Fregean theory of sense must be incompatible with externalism about content, since the latter is plausibly taken to entail that content is not transparent.

²¹ Frege (1979: 209-10).

Frege suggests that it is possible to doubt the truth of a correct analysis of a simple term only to the extent that one is not perfectly clear on its sense:

How is it possible, one may ask, that it should be doubtful whether a simple sign has the same sense as a complex expression if we know not only the sense of the simple sign but can recognize the sense of the complex one from the way it is put together? The fact is that if we really do have a clear grasp of the sense of the simple sign, then it cannot be doubtful whether it agrees with the sense of the complex expression. If this is open to question [...] then the reason must lie in the fact that we do not have a clear grasp of the sense of the simple sign.²²

But, crucially, this lack of clarity *is not* an impediment to ordinary competence, or even considerable expertise, with the term. This is clear, for Frege attributes such lack of clarity concerning the sense of ‘number’ to prominent mathematicians of his time. These mathematicians all attach the same sense to ‘number’, but they are in no position to know this fact, since they propound incompatible analyses of that sense.

More precisely, Frege’s argument is the following. The mathematicians’ lack of the capacity to re-identify the sense of ‘number’ is manifested in their inability to either give a correct analysis of it, or to recognize the correct analysis as such when they see it. But this inability does not prevent them from doing important mathematical work.²³ Thus, the lack of clarity Frege accuses them of does not entail that they do not grasp the sense of ‘number’ — in fact, he would probably have admitted that no one else (other than himself, of course) had a better grasp of that sense.

Moreover, it is not the case that each of the mathematicians assigns *a different* sense to ‘number’. For if that were the case, then:

The arithmetics of these mathematicians must be quite different. A sentence from the first mathematician must have a quite different sense from the equivalent-sounding sentence of the second mathematician.²⁴

²² Frege (1979: 211).

²³ Frege (1979: 222).

²⁴ Frege (1979: 215).

Frege presents this as a consequence that is patently absurd. Arithmetic is but one science, and therefore all of the mathematicians attach the same sense to ‘number’ — even though they do not know it:

Or is not the explanation rather that we have really to do with the same science; that this man *does* attach the same sense to the word ‘number’ as that man, only he doesn’t manage to get hold of it properly? Perhaps the sense appears to both of them through such a haze that when they make to get hold of it, they miss it.²⁵

According to Frege, therefore, the thoughts these mathematicians think are not transparent to any of them. The identity of their thoughts is fixed independently of their own opinions on the matter, by the unity of their science itself. This makes it possible for them to hold attitudes which are, in fact, inconsistent, without being in a position to recognize the inconsistency. Given all this context, it is natural to withhold the charge of incoherence or irrationality in these cases. This is a challenge to the notion of the rational availability of combinations of attitudes to a subject we need to address.²⁶

3. It is clear that we need to impose some sort of further constraint on the epistemic situation of the subject, if we are going to draw conclusions concerning the identity of her thoughts based on which combinations of attitudes are rationally available to her. In order to see how this would work, we can focus on trying to revise Evans’s Intuitive Criterion of Difference. As explained above, Evans’s principle is not fully satisfying in any case, as it only gives us a sufficient condition of difference and cannot be straightforwardly strengthened so as to give us a necessary one as well. Nevertheless, since the points I want to make here do not rest on a necessary condition of difference, in the rest of

²⁵ Frege (1979: 217).

²⁶ Cf. Burge (1990). Although Burge throughout insists on Frege’s rejection of the transparency of sense, he does not seem to recognize the consequences of this point for the Intuitive Criterion of Difference, which he appears to endorse in a footnote. Remarking that correct analyses appear to the mathematician as discoveries, and so can be doubted by one who does not doubt the corresponding self-identities, he remarks: “so by Frege’s test for the identity of senses, the senses of the explicans and the explicandum would be different.” (Burge (1992: 265, fn. 16)) But as Frege says about his own case (in the passage quoted in the text), one who has a fully clear grasp of the sense cannot rationally doubt the correct analysis, precisely because the two senses flanking the biconditional are identical. Thus the Intuitive Criterion only gives the correct result in the case where the subject does not suffer from such lack of clarity.

this section I will be using the Intuitive Criterion of Difference as my target for revisions. The same considerations would apply directly to the identification of thought-components and modes of combination.

Returning to (12) and (13), one obvious move would be to require that the subject not only understand the sentences and be rational, but further that her grasp of the sense be clear enough to make it possible for her to re-identify it on every occasion it is occurrently present in her consciousness. The revised principle, formulated in terms borrowed from Evans's original formulation, would look something like the following:

A thought *T* must be different from a thought *T'* if it is possible for someone whose understanding of *T* and *T'* satisfies a suitably high standard to rationally take contrasting attitudes towards them at the same time.

Here, of course, the 'suitably high standard' of understanding must be sufficient to avoid opacity.

As Frege's discussion shows, satisfying this high standard of understanding is not an easy matter: if even leading mathematicians lack this ability with respect to central terms of their science, ordinary understanding certainly does not guarantee that the criterion applies. This, however, gives us no grounds on which to object to it, since our search was not for a principle easily applicable in practice, but rather for a principle that would show us what to make of the very idea of Fregean thoughts.

Nevertheless, there is good reason to object to this formulation. The reason is that it leaves it quite unclear what the required standard of understanding consists in, or how we could recognize — even *in principle*, if not in practice — whether a person satisfies it or not. In one sense, of course, the answer to either of these questions is easy: one satisfies this standard with respect to a thought *T* if one is capable of identifying it whenever it is occurrently present in her consciousness. But we cannot be satisfied with this answer, for if this is the only grip we have on the relevant kind of understanding, our criterion will end up saying little more than just that two thoughts are distinct if

and only if someone with the ability to tell them apart can tell them apart. This is obviously true, but it provides no insight into the nature of thoughts.

Thus if we have no independent grip on what it takes for a subject to avoid opacity in a certain case, we also have no non-vacuous account of the identity of the corresponding thoughts. On the other hand, the prospects of giving a general non-trivial account of the kind of understanding required in order to avoid opacity seem extremely dim: the sheer diversity of intuitive examples of opacity makes it doubtful that there even *is* such a thing as a non-trivially specifiable state of understanding such that, for every thought whatsoever, opacity is avoided only if one is in it. For example, a different case of opacity that seems to come up in Frege concerns thoughts expressed by sentences containing temporal indexicals. Frege claims that a thought I expressed yesterday using “today” may be expressed today using “yesterday”. Yet it is clear that, if I have lost track of time, I might fail to keep track of the identity of that thought too.²⁷ But the lack of clarity I suffer from in *this* case is obviously very different from the lack of clarity that was behind the opacity in the case of Frege’s mathematicians. Why should we suppose that there is a non-trivially specifiable standard of clarity that will be sufficient to avoid *all* cases of opacity? This problem shows, I believe, that the revised version of the Intuitive Criterion of Difference is not acceptable.

The difficulty can, however, be avoided if we broaden our view to encompass not just an individual subject and her attitudes, but rather the whole of the practice she participates in and the relevant features of her environment. If we take this wider view we can avoid the need for a general, non-trivial specification of what it takes to avoid opacity. This move is actually made by Frege himself in his account of the disputing mathematicians: his argument was that opacity infected the mathematicians’ grasp of the sense of ‘number’, precisely because it was clear that all of the disputing mathematicians were practitioners of the same science. What sense each of them attached

²⁷ Frege (1997c: 331-332). See also Evans (2002b); McDowell (1998c).

to ‘number’ was not to be discovered by examining each individual mathematician in isolation, but rather by taking them collectively as participants in the science of mathematics.

Looking back at the ‘furze’ and ‘gorse’ example, it is easy to see that here too the problem lies in the way in which our criteria so far have been restricted to specifying conditions on the attitudes of an individual subject. What makes it the case that ‘furze’ and ‘gorse’ express the same sense, on the contrary, are facts about communal usage and about the interactions the community has had with the relevant plants — not facts about the dispositions of any individual subject considered in isolation. Once such communal facts are taken into account, it becomes clear that a subject who takes incompatible attitudes to (12) and (13) above is either suffering from opacity, or else is being irrational. This is because (as we are assuming, at least) there are no relevant differences in the community’s practice with the two words, nor do they derive from systematically different interactions with the plant.

Frege himself never discusses the possibility that the identity of thoughts may depend on facts about the physical environment. In fact, it is commonly assumed that his account of sense is incompatible with this possibility. Nevertheless, what he explicitly says about the sense of ‘number’ suggests he was well aware that sense is in general not fixed by any individual’s opinions or dispositions. For this reason, I think he could have easily accepted the dependence of sense on the physical environment as well.²⁸ In any case, the point here is not to remain faithful to Frege’s letter, but rather to sketch an account of content in Fregean spirit.

The crucial point that emerges from the discussion of cases of opacity, therefore, is that the standpoint from which the identity of thoughts is to be settled is not necessarily the standpoint of the subject who thinks them. The reason is, as we have seen, that the identity of thoughts depends

²⁸ This view has been explicitly argued for by Evans (1982; 2002b) and McDowell (1998b; 1998c). For arguments against it see Burge (1979). Frege’s own brief remarks on indexicals in Frege (1997c), however, indicate that he was aware of such context-dependent thoughts, and believed they did not pose a fundamental problem to the notion of sense.

on all kinds of contextual factors to which that subject need have no privileged insight. To credit the subject with the ability to re-identify her thoughts is often to credit her with knowledge that goes far beyond ordinary competence, or even expertise, in the relevant areas. Thus the standpoint from which the identity of thoughts is to be settled should rather be that of a Davidsonian interpreter: the standpoint of a subject who rationally interprets another subject, equipped with sufficient access to all of the relevant contextual facts.²⁹ As we saw, such an interpreter will sometimes find that attributing to her subjects some degree of ignorance about the identity of their own thoughts enables her to make better overall sense of them.

Incorporating these considerations, the criterion would look something like the following:

A thought *T* must be different from a thought *T'* if, given full knowledge of all the relevant facts about a context *C*, it is possible to interpret a subject *S* in *C* by attributing to *S* mutually exclusive attitudes towards *T* and *T'* without thereby attributing to *S* lack of understanding, failure of transparency, or incoherence.³⁰

²⁹ Here I am assuming the interpreter's contextual knowledge to include mastery of the interpreted subject's language; so the interpretation envisaged is not radical in Davidson's sense (see Davidson (1973)).

³⁰ In §2, I objected to the linguistic criterion for the individuation of senses (the Intersubstitutability Criterion), on the grounds that it made it too easy to attribute distinct senses to co-referential terms. One might wonder whether the new criterion, incorporating as it does the possibility of non-transparent senses, might not make it too hard: what reasons could we have to assign distinct senses to co-referential words, if we can always avoid doing so by attributing instead opacity to the subjects who use them? I do not think this worry is justified, however. In practice, the decision to make distinctions between senses will face the decision to attribute opacity in an interpretative tradeoff of a kind we are in any case used to.

We can consider some examples to see how such tradeoffs might be resolved. To begin with, consider once again the case of Jones, our botanically ignorant subject. Although Jones has used the word 'gorse' all of her life, she only recently picked up the word 'furze' in conversation, and all she knows about it is that it refers to a plant that grows in dry areas. Thus, it is intuitive to take her to defer in her use of 'furze' to people who understand it better than she does. We should not, therefore, attribute to her an idiosyncratic sense for 'furze'; it is much more plausible to attribute to her opacity about its sense. Now consider the case of 'Phosphorus' and 'Hesperus' for contrast. Consider, more specifically, the usage of an astronomer prior to the discovery that the appearance of a bright star visible in the morning and of a bright star visible in the evening are both due to the same planet. We have, in this case, no reason to attribute to our subject any lack of clarity concerning the senses of the two words. Thus we should agree with Frege's judgment that our subject does not use 'Hesperus' and 'Phosphorus' with the same sense. Of course other cases will not be as simple as these toy examples; but this is not to say that we cannot, in general, make principled decisions concerning the identity of senses.

By “context” here I mean the totality of social and environmental facts that may be relevant, including facts about the nature of the subject’s practices she and every other actual member of these practices may be ignorant of.³¹

4. The account just sketched naturally meshes with the holistic constraint on concept-possession that was suggested in Chapter 2 — in fact, it provides a rationale for it, grounded in the nature of Fregean thoughts. I will close this Appendix by revisiting the example discussed in that chapter, in light of the present account.

On the present account, part of what fixes the identity of the concept of negation is that there are some characteristic relations of incompatibility among attitudes in which it figures — in particular (assuming, once again, classical logic to be correct) attitudes of acceptance towards any thought of the form ‘ $\sim\sim A$ ’ are incompatible with rejecting or being agnostic about ‘ A ’. Other things being equal, if one believes ‘ $\sim\sim A$ ’ one should not reject or be agnostic about ‘ A ’ (unless one is also thereby led to reconsider ‘ $\sim\sim A$ ’, of course).

Now, as we have seen, this is not to say that on every occasion a rational subject who understands the concept of negation will be disposed to conform her thinking to this principle. For one thing, in a particular context it might be better to attribute to the subject opacity concerning the thought that replaces ‘ A ’ in our schema, rather than either a misunderstanding of negation or irrationality. But cases of genuinely *logical* error can be accommodated by the present view as well, without the attribution of misunderstanding or irrationality.

³¹ It follows that there will, in general, be a limit to how much we can know about the identity or distinctness of thoughts available in a given context. This limit is not *ad hoc*, but simply reflects the fact that we often — perhaps always — do not have full knowledge of all the relevant aspects of the context. The existence of such a limit does not deprive our criterion of a point, of course: the point was to say something informative about the identity and distinctness of thoughts, not to provide a decision procedure capable of making perfectly sharp distinctions in practice.

To return to the example I have been using throughout, a non-classical logician may fail — because of opacity — to recognize in a classical logician's explicitly stated rules of inference her own concept of negation (she may even fail to recognize in them a coherent concept at all). In such a case, we may still attribute the concept of negation to her. We would do this not because we would hypothesize that she must possess some hidden or suppressed disposition to infer in accordance with the rules of inference she explicitly rejects, but rather because her overall outlook is that of a competent enough member of our conceptual practice. What that would require, of course, is her continuing reliability in enough of the rest of the norms of that practice.

BIBLIOGRAPHY

- ANSCOMBE, G. E. M. (2000): *Intention*. (Cambridge, MA: Harvard University Press.)
- BEALER, G. (1999): 'A Theory of the A Priori'. (*Noûs*, vol. 33, *Supplement: Epistemology*: 29-55.)
- BOGHOSSIAN, P. (1996): 'Analyticity Reconsidered'. (*Noûs*, vol. 30: 360-391.)
- BOGHOSSIAN, P. (2000): 'Knowledge of Logic'. (In Boghossian, P., and Peacocke, C., eds., *New Essays on the A Priori* (229-254). New York, NY: Oxford University Press.)
- BOGHOSSIAN, P. (2003): 'Blind Reasoning (I)'. (*Aristotelian Society Supplementary Volumes*, vol. 77: 225-248.)
- BONJOUR, L. (1998): *In Defense of Pure Reason*. (New York, NY: Cambridge University Press.)
- BONJOUR, L. and SOSA, E. (2003): *Epistemic Justification: Internalism vs. Externalism, Foundations vs. Virtues*. (Malden, MA: Blackwell Publishing.)
- BURGE, T. (1978): 'Belief and Synonymy'. (*The Journal of Philosophy*, vol. 75 (3): 119-138.)
- BURGE, T. (1979): 'Sinning Against Frege'. (In *Truth, Thought, Reason: Essays on Frege* (213-241). New York, NY: Clarendon Press-Oxford University Press, 2005.)
- BURGE, T. (1990): 'Frege on Sense and Linguistic Meaning'. (In *Truth, Thought, Reason: Essays on Frege* (242-270). New York, NY: Clarendon Press-Oxford University Press, 2005.)
- BURGE, T. (1992): 'Frege on Knowing the Third Realm'. (In *Truth, Thought, Reason: Essays on Frege* (299-316). New York, NY: Clarendon Press-Oxford University Press, 2005.)
- CHALMERS, D. (1996): *The Conscious Mind: in Search of a Fundamental Theory*. (New York: Oxford University Press.)
- CHALMERS, D. AND JACKSON, F. (2001): 'Conceptual Analysis and Reductive Explanation'. (*Philosophical Review* 110: 315-61.)
- CHALMERS, D. (2002a): 'Does Conceivability Entail Possibility?'. (In T. Gendler and J. Hawthorne, eds., *Conceivability and Possibility*, (145-221). New York: Oxford University Press.)

- CHALMERS, D. (2002b): 'On Sense and Intension'. (Retrieved May 9th 2007 from: <<http://consc.net/papers/intension.pdf>>.)
- CHALMERS, D. (2002c): 'The Components of Content'. (Retrieved May 9th 2007 from: <<http://consc.net/papers/content.pdf>>.)
- CHALMERS, D. (2006): 'The Foundations of Two-Dimensional Semantics'. (In M. Garcia-Carpintero and J. Macia, eds., *Two-Dimensional Semantics: Foundations and Applications*. New York, NY: Oxford University Press. Retrieved May 9th 2007 from <<http://consc.net/papers/foundations.html>>.)
- COFFA, A. (1993): *The Semantic Tradition from Kant to Carnap*. (Cambridge: Cambridge University Press.)
- COHEN, S. (2002): 'Basic Knowledge and the Problem of Easy Knowledge'. (*Philosophy and Phenomenological Research*, vol. LXV: 309-329.)
- DAVIDSON, D. (1973): 'Radical Interpretation'. (In Davidson, D., *Inquiries into Truth and Interpretation* (125-140). New York, NY: Oxford University Press, 1984.)
- DAVIDSON, D. (1974): 'On the Very Idea of a Conceptual Scheme'. (In Davidson, D., *Inquiries into Truth and Interpretation* (183-198). New York, NY: Oxford University Press, 1984.)
- DAVIDSON, D. (1983): 'A Coherence Theory of Truth and Knowledge'. (In Davidson, D., *Subjective, Intersubjective, Objective* (137-153). New York, NY: Oxford University Press, 2001.)
- DAVIES, M. (2004): 'On Epistemic Entitlement (II): Epistemic Entitlement, Warrant Transmission and Easy Knowledge'. (*Aristotelian Society, Supplementary Volumes*, vol. 79: 213-244.)
- DUMMETT, M. (1975): 'Frege's Distinction Between Sense and Reference'. (In *Truth and Other Enigmas* (116-144). Cambridge, MA: Harvard University Press, 1978.)
- DUMMETT, M. (1981): *Frege: Philosophy of Language*. (London: Duckworth.)
- EVANS, G. (1982): *The Varieties of Reference*. (J. McDowell, ed., New York, NY: Clarendon Press-Oxford University Press.)
- EVANS, G. (2002a): 'Things Without the Mind'. (In *Collected Papers* (249-290). New York, NY: Clarendon Press-Oxford University Press.)
- EVANS, G. (2002b): 'Understanding Demonstratives'. (In *Collected Papers* (291-321). New York, NY: Clarendon Press-Oxford University Press.)
- FIELD, H. (2001): 'Apriority as an Evaluative Notion'. (In *Truth and the Absence of Fact* (361-391). New York, NY: Oxford University Press.)

- FREGE, G. (1967): *Begriffsschrift*. (In J. van Heijenoort, ed., *From Frege to Gödel: A Sourcebook in Mathematical Logic 1879-1931* (1-82). Cambridge, MA: Harvard University Press.)
- FREGE, G. (1979): 'Logic in Mathematics'. (In *Posthumous Writings* (203-250). New York, NY: Blackwell Publishers.)
- FREGE, G. (1980): *The Foundations of Arithmetic*. (J. L. Austin, trans. Evanston, IL: Northwestern University Press.)
- FREGE, G. (1997a): *Grundgesetze der Arithmetik, vol. I*. (In M. Beaney, ed., *The Frege Reader* (194-223). New York, NY: Blackwell Publishers.)
- FREGE, G. (1997b): 'On *Sinn* and *Bedeutung*'. (In M. Beaney, ed., *The Frege Reader* (151-171). New York, NY: Blackwell Publishers.)
- FREGE, G. (1997c): 'Thought'. (In M. Beaney, ed., *The Frege Reader* (325-345). New York, NY: Blackwell Publishers.)
- FREGE, G. (1997d): 'Letter to Jourdain, Jan. 1914'. (In M. Beaney, ed., *The Frege Reader* (319-321). New York, NY: Blackwell Publishers.)
- FREGE, G. (1997e): 'Function and Concept'. (In M. Beaney, ed., *The Frege Reader* (130-148). New York, NY: Blackwell Publishers.)
- HYLTON, P. (2002): *Russell, Idealism and the Emergence of Analytic Philosophy*. (New York, NY: Clarendon Press-Oxford University Press.)
- JACKSON, F. (1998): *From Metaphysics to Ethics: A Defense of Conceptual Analysis*. (New York: Oxford University Press.)
- JACKSON, F. (2000): 'Representation, Skepticism, and the A Priori'. (In P. Boghossian, and C. Peacocke, eds., *New Essays on the A Priori* (320-332). New York, NY: Oxford University Press.)
- KANT, I. (1998): *Critique of Pure Reason*. (Paul Guyer and Allen Wood, trans. New York, NY: Cambridge University Press.)
- KAPLAN, D. (1989): 'Demonstratives: An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives'. (In J. Almog, J. Perry, H. Wettstein, eds., *Themes from Kaplan* (481-564). New York, NY: Oxford University Press.)
- KRIPKE, S. (1979): 'A Puzzle about Belief'. (In A. P. Martinich, ed., *The Philosophy of Language* (405-431). New York, NY: Oxford University Press, 2001.)
- KRIPKE, S. (1982): *Wittgenstein on Rules and Private Language*. (Cambridge, MA: Harvard University Press.)
- KRIPKE, S. (1999): *Naming and Necessity*. (Cambridge, MA: Harvard University Press.)

- LEVINE, J. (1998): 'Acquaintance, Denoting Concepts, and Sense'. (*The Philosophical Review*, Vol. 107 (3): 415-445.)
- LEWIS, D. (1999): 'Putnam's Paradox'. (In *Papers in Metaphysics and Epistemology* (56-77). New York, NY: Cambridge University Press.)
- MCDOWELL, J. (1994): *Mind and World*. (Cambridge, MA: Harvard University Press.)
- MCDOWELL, J. (1998a): 'Knowledge and the Internal'. (In *Meaning, Knowledge and Reality* (395-413). Cambridge, MA: Harvard University Press.)
- MCDOWELL, J. (1998b): 'On the Sense and Reference of a Proper Name'. (In *Meaning, Knowledge and Reality* (171-198). Cambridge, MA: Harvard University Press.)
- MCDOWELL, J. (1998c): 'De Re Senses'. (In *Meaning, Knowledge and Reality* (214-227). Cambridge, MA: Harvard University Press.)
- MCDOWELL, J. (1998d): 'Criteria, Defeasibility and Knowledge'. (In *Meaning, Knowledge and Reality* (214-227). Cambridge, MA: Harvard University Press.)
- MCDOWELL, J. (2005): 'Evans's Frege'. (In J. L. Bermúdez, ed., *Thought, Reference and Experience* (42-65). New York, NY: Oxford University Press.)
- MOORE, G. (1899): 'The Nature of Judgment'. (*Mind*, Vol. 8 (30): 176-193).
- PEACOCKE, C. (1992): *A Study of Concepts*. (Cambridge, MA: Massachusetts Institute of Technology Press.)
- PEACOCKE, C. (2000): 'Explaining the A Priori: the Programme of Moderate Rationalism'. (In P. Boghossian, and C. Peacocke, eds., *New Essays on the A Priori* (255-285). New York, NY: Oxford University Press.)
- PEACOCKE, C. (2003): 'Implicit Conceptions, Understanding and Rationality'. (In B. Ramberg, ed., *Reflections and Replies: Essays on the Philosophy of Tyler Burge* (117-152). Cambridge, MA: Massachusetts Institute of Technology Press.)
- PEACOCKE, C. (2004): *The Realm of Reason*. (New York, NY: Oxford University Press.)
- PRYOR, J. (2000): 'The Skeptic and the Dogmatist'. (*Noûs*, vol. 34: 517-549.)
- PRYOR, J. (2004): 'What's Wrong with Moore's Argument?'. (*Philosophical Issues 14: Epistemology*: 349-378.)
- PUTNAM, H. (1975): 'The Meaning of "Meaning"'. (In K. Gudenson, ed., *Language, Mind, and Knowledge* (131-91). Minneapolis: Minnesota Studies in the Philosophy of Science-University of Minnesota Press.)

- QUINE, W. V. O. (1935): 'Truth by Convention'. (In *The ways of Paradox* (77-106). Cambridge, MA: Harvard University Press.)
- QUINE, W. V. O. (1954): 'Carnap and Logical Truth'. (In *The ways of Paradox* (107-132). Cambridge, MA: Harvard University Press.)
- QUINE, W. V. O. (1980): 'Two Dogmas of Empiricism'. (In *From a Logical Point of View* (20-46). Cambridge, MA: Harvard University Press.)
- RICKETTS, T. (2001): 'Truth and Propositional Unity in Early Russell'. (In J. Floyd, ed., *Future Pasts: The Analytic Tradition in Twentieth-Century Philosophy* (101-121). New York: NY: Oxford University Press.)
- RICKETTS, T. (1986): 'Objectivity and Objecthood: Frege's Metaphysics of Judgment'. (In L. Haaparanta and J. Hintikka, eds., *Frege Synthesized* (65-96). Dordrecht: D. Reidel Publishing Company.)
- RICKETTS, T. (1996): 'Logic and Truth in Frege (I)'. (*Proceedings of the Aristotelian Society, Supplementary Volumes*, vol. 70: 121-175.)
- RICKETTS, T. (2002): 'Wittgenstein against Frege and Russell'. (In E. Reck, *From Frege to Wittgenstein: Perspectives in Early Analytic Philosophy* (228-251). New York, NY: Oxford University Press.)
- RICKETTS, T. (forthcoming): 'Concepts, Objects, and the Context Principle'. (Forthcoming in T. Ricketts, ed., *The Cambridge Companion to Frege*. Cambridge University Press.)
- RÖDL, S. (2007): *Self-Consciousness*. (Cambridge, MA: Harvard University Press.)
- RUSSELL, B. (1905): 'On Denoting'. (*Mind*, vol. 14 (56): 479-493.)
- RUSSELL, B. (1906): 'On the Nature of Truth'. (*Mind*, vol. 15 (60): 528-533.)
- RUSSELL, B. (1996): *The Principles of Mathematics*. (New York, NY: W. W. Norton and Company.)
- RUSSELL, B. (1997): *The Problems of Philosophy*. (New York, NY: Oxford Paperbacks-Oxford University Press.)
- STRAWSON, P. (1999): *The Bounds of Sense*. (New York, NY: Routledge.)
- STRAWSON, P. (2002): *Individuals*. (New York, NY: Routledge.)
- THOMPSON, M. (1995): 'The Representation of Life'. (In R. Hursthouse, ed, *Virtues and Reasons: Philippa Foot and Moral Theory* (247-296). New York: Oxford University Press.)
- WILLIAMSON, T. (2000): *Knowledge and Its Limits*. (New York, NY: Oxford University Press.)
- WILLIAMSON, T. (2003): 'Blind Reasoning (II): Understanding and Inference'. (*Aristotelian Society, Supplementary Volumes*, vol. 77: 249-293.)

- WILLIAMSON, T. (2006): 'Conceptual Truth'. (*Aristotelian Society, Supplementary Volumes*, vol. 80: 1-41.)
- WITTGENSTEIN, L. (1958): *Philosophical Investigations*. (G. E. M. Anscombe, trans. Englewood Cliffs, NJ: Prentice Hall.)
- WRIGHT, C. (1983): *Frege's Conception of Numbers as Objects*. (Aberdeen: Aberdeen University Press.)
- WRIGHT, C. (1985): 'Facts and Certainty'. (*Proceedings of the British Academy*, 71: 429-472.)
- WRIGHT, C. (2003): 'Some Reflections on the Acquisition of Warrant by Inference'. (In S. Nuccetelli, ed., *New Essays on Semantic Externalism and Self-Knowledge* (57-78). Cambridge, MA: Massachusetts Institute of Technology Press.)
- WRIGHT, C. (2004): 'On Epistemic Warrant (I): Warrant for nothing (and foundations for free)?' (*Aristotelian Society, Supplementary Volumes*, vol. 79: 168-211.)
- YABLO, S. (2002): 'Coulda, Woulda, Shoulda'. (In T. Gendler and J. Hawthorne, eds., *Conceivability and Possibility* (441-492), New York, NY: Oxford University Press.)